Correcting deficiencies in approximate density functionals

Jonathan David Gledhill

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

2015

Correcting deficiencies in approximate density functionals

Jonathan David Gledhill

Abstract

In the last fifty years, approximate density functional theory (DFT) has become firmly established as the *de facto* standard for electronic structure calculations in chemistry. Although the theory itself is formally exact, approximations must be made for the unknown *exchange–correlation* (XC) functional, and whilst many successful approximate functionals exist, a number of deficiencies still persist, leading to many cases where the approximation breaks down completely. This thesis addresses two prevalent deficiencies, and examines some novel approaches to reducing and eliminating them.

Chapter 1 provides a background to electronic structure theory, with particular reference to the approximate solution of the electronic Schrödinger equation through *ab initio* wavefunction methods. Chapter 2 then provides the formal justification for DFT as an alternative to wavefunction-based methods, and outlines common approximations to the XC functional. Two prominent deficiencies of approximate DFT are discussed: delocalisation error due to non-linearity in the energy variation with number of electrons, and incorrect long-range behaviour of the XC potential.

Chapter 3 examines a system-dependent tuning technique for the range-separated hybrid class of XC functionals, whereby the range-separation parameter is non-empirically tuned to self-consistent energy-linearity conditions, which has been successfully used to improve the calculation of quantities affected by the delocalisation error. A full, systematic assessment of this tuning technique is provided, and it is demonstrated that the success of the technique is aided by a convenient cancellation of errors.

In Chapter 4, the tuned functionals are applied to quantities relevant to conceptual DFT. It is shown that functionals tuned to the energy conditions of Chapter 3 remain appropriate for calculation of the electronegativity from orbital energies, however the density variation with number of electrons—described by the Fukui function—is better modelled by conventional non-tuned functionals.

Finally, an entirely new approach to functional development is provided in Chapter 5. The behaviour of a functional under density scaling is used to impose homogeneity constraints on a simple functional form, culminating in an *electron-deficient* functional that satisfies the appropriate energy-linearity condition and exhibits the correct asymptotic XC potential.

Author's Declaration

The material contained within this thesis has not previously been submitted for a degree at the University of Durham or any other university. The research reported within this thesis has been conducted by the author unless otherwise indicated.

Copyright Notice

Contents

Contents **i**

List of Figures **iii**

List of Tables **v**

Selected definitions **vii**

Acknowledgements **ix**

1 Quantum Mechanics **1**
1.1 The Schrödinger equation 2
1.2 Hartree–Fock theory 6
1.3 Correlated methods 12

2 Density functional theory **19**
2.1 Thomas–Fermi theory 20
2.2 The Hohenberg–Kohn theorems 21
2.3 The Levy constrained search approach 25
2.4 Kohn–Sham theory 27
2.5 Exchange–correlation functionals 29
2.6 Time-dependent DFT 36
2.7 Specific failures of conventional DFT 40

3 Tuned range-separated hybrid functionals **49**
3.1 Functional forms 51
3.2 Tuning functionals 53
3.3 Preliminary assessments 58
3.4 Assessment of tuning methods 69
3.5 Conclusions 83

4 Fukui function from tuned functionals **87**
4.1 Conceptual DFT 87
4.2 Near-linearity of the density gradient 91
4.3 Functional tuning 93
4.4 Conclusions and further considerations 101

5 Development of novel explicit density functionals **103**
5.1 Scaling relations 105
5.2 Combining Coulomb and XC terms 108
5.3 Parr–Ghosh functional 117
5.4 Development of an electron-deficient functional 119
5.5 Summary of the ED functional scheme 131
5.6 Analysis and performance of the ED functionals 131
5.7 Conclusions 142

6 Conclusions **145**

A ED Exchange–correlation kernel **151**

B Presented work **155**
Publications 155
Meetings attended 156

Bibliography **157**

List of Figures

Chapter 3

3.1 Schematic illustration of the partitioning of DFT and exact-exchange in hybrid and range-separated functionals 52

3.2 Schematic representation of the quantities involved in the tuning methods 56

3.3 Variation of $\|J_{\mathrm{H,H}}\|_1$ and $\|J_{\mathrm{H,L}}\|_1$ with μ for various RSH functional forms 60

3.4 Deviation of calculated I^{N_0} and A^{N_0} from experimental $I_0^{N_0}$ and $A_0^{N_0}$ for various RSH functional forms 62

3.5 The degree to which the HOMO Koopmans condition is satisfied by LC-B3LYP, for successive ionisations of carbon, as a function of μ 63

3.6 *E vs N* deviation curves for successive pairs of carbon species, using tuned LC-B3LYP 65

3.7 The degree to which the HOMO Koopmans condition is satisfied by correlation-free LC-B3LYP, for successive ionisations of carbon, as a function of μ 67

3.8 *E vs N* deviation curves for successive pairs of carbon species, using correlation-free LC-B3LYP at very large μ 67

3.9 Error in calculated (ΔSCF) ionisation energies of carbon species, as a function of μ, for LC-B3LYP 68

3.10 HOMO–LUMO gap, and its deviation from experimental fundamental gap, calculated using tuned BNL 72

3.11 HOMO–LUMO gap, and its deviation from experimental fundamental gap, calculated using tuned LC-B3LYP 73

3.12 Schematic relationship between an *E vs N* curve and an *E vs N* deviation curve. 79

3.13 *E vs N* deviation curves and exact slopes for the carbon atom using RHS tuning norms 80
3.14 *E vs N* deviation curves and exact slopes for the carbon atom using LHS tuning norms 81
3.15 *E vs N* deviation curves and exact slopes for the carbon atom using double-segment tuning norms 82

Chapter 4

4.1 Example Fukui difference plots for the carbon atom 93
4.2 Fukui functions for the carbon atom, computed using the finite difference method from integer-electron species 95
4.3 Fractional Fukui functions for the carbon atom, computed using the generalised finite difference method, equations (4.24) and (4.25), with $x = \pm 0.01$ 96
4.4 Fukui difference plots for the carbon atom, using LC-B3LYP with a variety of μ values 98
4.5 Expansion of the chemically significant region of the carbon Fukui difference plots, around the van der Waals radius 99

Chapter 5

5.1 Comparison of exact k_Ω values from equation (5.16) with approximate values from equation (5.19) 110
5.2 Errors in $E_{\text{XC}}[\rho]$, compared to near-exact ZMP values, for a range of $B[\rho]$ functionals and unmodified Fermi–Amaldi 112
5.3 k_Ω values for $B[\rho] = \alpha N^p$ using different optimisation criteria, compared with near-exact values 115
5.4 Errors in $E_{\text{XC}}[\rho]$ with respect to near-exact ZMP values, for $B[\rho] = \alpha N^p$ using different optimisation criteria, along with unmodified Fermi–Amaldi 116
5.5 Linear-regression fits of the ED approximate energy functionals against near-exact values 122
5.6 Comparison of approximations for k_F and k_G from equations (5.54) and (5.55) with their exact equivalents from Table 5.4 128
5.7 Effective homogeneities of the exchange–correlation functional from self-consistent ED1 and ED2 calculations, compared with near-exact k_{XC}^- values from Ref. 244 134
5.8 Exchange–correlation potentials along the bond axis for CO and PN, compared to the near-exact ZMP potential from Ref. 257 140

List of Tables

Chapter 1

1.1 Definitions of fundamental and derived atomic units 4

Chapter 3

3.1 Summary of the 22 tuning norms assessed in Chapter 3 59

3.2 Optimal range-separation parameters μ^* that minimise H for successive ionisations of carbon using LC-B3LYP 64

3.3 Optimal range-separation parameters μ^* that minimise the error in $I^{\Delta\text{SCF}}$ for successive ionisations of carbon using LC-B3LYP 69

3.4 Mean absolute deviations, from exact reference values, of various key quantities computed using conventional functionals and each tuning norm 75

3.5 Deviations (in eV), from exact reference values, of various key quantities computed for the carbon atom using selected tuning norms 78

Chapter 4

4.1 Errors in the electronegativity χ for conventional and tuned functionals 100

Chapter 5

5.1 Calculated prefactors α computed using $B[\rho]$ in equation (5.26), along with mean absolute errors (MAE) and sums of square errors (SSE) in the resulting $E_{\text{XC}}[\rho]$ values, compared to near-exact ZMP values 111

5.2 Optimised parameters for the functional $B[\rho] = \alpha N^p$, along with the errors in k_Ω and E_{XC}, for various minimisation conditions 114

5.3 Deviation of the effective homogeneities of the Parr–Ghosh (PG) functional from near-exact k_{XC}^{-}, along with the errors in total energy (from near-exact values) quoted in Ref. 250 119
5.4 Exact effective homogeneities k_F and k_G required for functionals $F[\rho]$ and $G[\rho]$, in order to ensure $k_{\mathrm{XC}} = k_{\mathrm{XC}}^{-}$, for equations (5.37) and (5.38) 121
5.5 Parameters defining the "optimal" ED1 and ED2 functionals, as determined from linear regression fits 123
5.6 Representative Hellmann–Feynman tests for CO, using the ED2 functional 125
5.7 Mean absolute errors in the exchange–correlation and HOMO energies, for self-consistent calculations using ED1 and ED2 126
5.8 Optimal values of m based on self-consistent calculations with ED1 and ED2, using a system-dependent prefactor 130
5.9 Parameters defining the ED1 functional 132
5.10 Parameters defining the ED2 functional 132
5.11 Exchange–correlation energies (in E_{h}), compared to near-exact values from Ref. 244 134
5.12 HOMO energies (in E_{h}), compared to the negative of the exact vertical ionisation potential from Ref. 149 135
5.13 HOMO energies (in E_{h}), compared to the negative of the RMP2 ionisation potential I 136
5.14 Vertical excitation energies (in eV), compared to experimental values from Ref. 191 137
5.15 Static isotropic polarisabilities (in a.u.), compared to reference BD(T) values from Ref. 94 139

Selected definitions

A	Electron affinity
AO	Atomic orbital
a.u.	Atomic units
aug-cc-pVTZ	Dunning correlation-consistent triple-zeta basis set with additional diffuse functions
BD	Brueckner doubles
BD(T)	Brueckner doubles with perturbative triples
BLYP	Becke's 1988 GGA exchange functional with LYP correlation
B3LYP	Becke 3-parameter global hybrid exchange–correlation functional
BNL	Baer–Neuhauser–Livshits RSH exchange–correlation functional
CAM	Coulomb attenuated method of range-separation, through equation (3.2), with attenuation parameters α, β, and μ
CC	Coupled cluster method
CCSD	Truncation of CC at the singles and doubles level
CI	Configuration interaction
CS: m	Homogeneity of degree m under coordinate scaling
DFT	Density functional theory
DS: k	Homogeneity of degree k under density scaling
ε_{H}	Energy of the HOMO
ε_{L}	Energy of the LUMO
E	Total (electronic) energy
ED	Electron-deficient, referring to a functional appropriate for this side of an integer-electron system
E *vs* N	Energy variation with respect to the number of electrons
E_{XC}	Exchange–correlation energy
E_{X}^{0}	Exact orbital exchange included in hybrid and range-separated functionals
FA	Fermi–Amaldi functional
FCI	Full configuration interaction, including all possible excited state determinants
FD	Finite difference
GGA	Generalised gradient approximation
GH	Global hybrid functional incorporating a fixed proportion of E_{X}^{0}
GKS	Generalised Kohn–Sham theory
GTO	Gaussian-type orbital

HF	Hartree–Fock theory
HK	Hohenberg–Kohn theorems
HOMO	Highest occupied molecular orbital
I	Ionisation potential
k	Homogeneity of a functional under density scaling
KS	Kohn–Sham formulation of DFT
LC	Long-range correction method of range-separation, through equation (3.1). Equivalent to CAM with $\alpha = 0$ and $\beta = 1$
LDA, LSDA	Local density approximation, local spin density approximation
LHS	Left-hand (electron-deficient) side of an integer-electron system
LUMO	Lowest unoccupied molecular orbital
LYP	Lee–Yang–Parr GGA correlation functional
μ^*	Optimal range-separation parameter for a given tuning method
m	Homogeneity of a functional under coordinate scaling
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MESIE	Many-electron self-interaction error
MPPT	Møller–Plesset perturbation theory
MP2	Second-order truncation of MPPT
N	Number of electrons
PBE	Perdew–Burke–Ernzerhof GGA exchange–correlation functional
PG	Parr–Ghosh functional
RHF	Restricted Hartree–Fock method
RHS	Right-hand (electron-abundant) side of an integer-electron system
ROHF	Restricted open-shell Hartree–Fock method; paired electrons occupy orbitals with identical spatial parts
RSH	Range-separated hybrid functional, incorporating a variable proportion of E_{x}^{0} determined by the inter-electron distance
SCF	Self-consistent field method of iterating a calculation to a converged minimum
ΔSCF	Shorthand for the energy difference between two successive integer-electron species, *i.e.* computed I or A value
SIE	Self-interaction error, synonymous with delocalisation error
SSE	Sum of square errors
TDKS	Time-dependent Kohn–Sham theory
TDDFT	Time-dependent density functional theory
TF	Thomas–Fermi approximations to the energy in terms of density functionals
TFD	Thomas–Fermi–Dirac correction to the TF scheme
UEG	Uniform electron gas
UHF	Unrestricted open-shell Hartree–Fock method; paired electrons occupy orbitals with spatial parts that are allowed to differ
XC	Exchange–correlation
ZMP	Near-exact Zhao–Morrison–Parr reference values

Acknowledgements

A work such as this could never be attributed to one person alone. Throughout my time in Durham I have had the pleasure of working with and learning from a great many individuals whose teaching, support and collaboration have been invaluable to the completion of my Ph.D.

First and foremost, my sincere thanks go to David Tozer, whose patient supervision, boundless enthusiasm and painstaking attention to detail have given me the freedom to explore and discover whilst keeping me grounded and focused. Never again will I hear the words "operator error"—delivered with such gusto—without travelling back to the long hours debugging Fortran, but the good humour and helpful suggestions that accompanied them were always an inspiration.

A great deal of credit must also be given to Michael Peach and Alex Borgoo for their invaluable advice and assistance, from teaching me basic principles and techniques through to collaborations on the RSH-tuning and homogeneity projects, respectively. I am indebted to them for always finding the time to offer support and encouragement (and coffee). Thanks must also go to Martin Walker, whose friendship, advice and laid-back attitude have been a constant antidote to the stresses—and reminder of the pleasures—of academic life. I also thank Frank De Proft for helpful discussions regarding the Fukui function, and the EPSRC for financial support.

Although too numerous to mention by name, my thanks go to all of the friends I have made over the years I have spent in Durham, in particular the denizens of 200 and 200X, for keeping me sane, solvent and above all motivated to keep going. Stephen, Paul and Andy must get special mention for helping me escape back into the outdoors, and without whom my experience and prospects would be very different. I also thank Pete, Dave and Emma for largely stress-free living, Caroline for mutual moral support, and Neal for his encouragement, reminding me to enjoy myself, and his truly selfless attitude.

Finally my eternal gratitude must go to my family, without whose support and belief over the years I would never have had the chance to be in this position. Thank you.

1

Quantum Mechanics

The foundations of quantum mechanics are briefly discussed, with particular reference to the Schrödinger equation as a determinative of molecular electronic structure. The importance of the wavefunction as a descriptor for a particular electronic state is then highlighted, and a discussion is presented of its use in methods to solve electronic structure problems in chemistry. Finally, a brief hierarchy of techniques used to approximately solve the Schrödinger equation is outlined, with reference to the inherent challenges that lead to the consideration of wavefunction-free alternatives.

The postulates of quantum mechanics,[1–3] culminating in the time-dependent Schrödinger equation, together describe the behaviour of all known microscopic matter. It is the approximate solution, through various methods, to this equation that drives electronic structure theory, in order to model and predict the structure and properties of molecular systems.

Quantum mechanics has a history stretching throughout the 20th century, beginning with early efforts to correct failures of traditional Newtonian mechanics. Planck's quantisation of electromagnetic radiation[4] began the development, followed by Einstein's rationalisation of the photoelectric effect[5] and Bohr's hypothesis of quantised angular momentum.[6] The links between electromagnetic radiation and matter were cemented with de Broglie's conjecture[7] that any moving body possesses a wavelength inversely proportional to its linear

momentum, giving rise to the concept of wave–particle duality.

Both Schrödinger[8–13] and Heisenberg[14] independently developed formulations of quantum mechanics. Whilst the two formalisms were developed and expressed using different mathematical concepts, the two are equivalent. Although the matrix terminology of Heisenberg's formalism is best suited to formal manipulations of the theory, the familiarity of Schrödinger's functions and differential calculus lends itself more readily to interpretation of the theory and to calculating numerical results, ultimately leading to its widespread fame.

1.1 The Schrödinger equation

The time-dependent Schrödinger equation is given by

$$i\hbar\frac{\partial\Psi}{\partial t} = \hat{H}\Psi\,, \tag{1.1}$$

where $\Psi = \Psi(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n, t)$ is the wavefunction, dependent on both time t and the coordinates $\boldsymbol{x}$ of each of the n particles in the system. Here, $\boldsymbol{x}$ is a coordinate combining the spatial position $\boldsymbol{r}$ and spin coordinate s of the particle. $\hat{H}$ is the Hamiltonian operator for the system, the terms of which describe the contributions to the total energy E of the system, and $\hbar$ is the reduced Planck constant $\hbar = h/2\pi$.

For many purposes we are not concerned with time-dependent interactions, and for applications where the potential energy is independent of time, the Schrödinger equation is separable into components describing the time and space variation of the wavefunction, respectively. The time-dependent component of such a wavefunction is a complex phase factor, $\mathrm{e}^{-iEt/\hbar}$, which does not affect the absolute square of the wavefunction. Given that only the absolute square of the wavefunction—rather than the wavefunction itself—has physical meaning[15] (corresponding to the probability density of finding a particle at a given point in space), the phase factor has no effect on the interpretation of the wavefunction and we can remove the direct time-dependence. The resulting time-*independent* Schrödinger equation is given by

$$\hat{H}\psi = E\psi\,, \tag{1.2}$$

where $\psi = \psi(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ is the time-independent wavefunction.

1.1.1 BORN–OPPENHEIMER APPROXIMATION

A further simplification we can make is the approximation outlined by Born and Oppenheimer,[16] which involves separating the Schrödinger equation into two further parts: electronic and nuclear. This allows us to solve separately an *electronic* Schrödinger equation—where the nuclei are regarded as fixed in position and exerting a static external potential on the electrons. Repeating this for a set of external potentials arising from different nuclear configurations then builds up a potential energy surface for the system.

This decoupling of the nuclear and electronic motions is made possible by the large difference in their masses: any change in nuclear position produces a near-instantaneous response in the electrons, and so from the point of view of the electrons, the nuclei can be regarded as fixed. The approximation relies on the assumption that electrons behave adiabatically with respect to nuclear motion, *i.e.* that the electronic state is unaffected by a change in the nuclear position. When electronic states are sufficiently separated in energy this assumption is generally valid, however as the energy of two or more states approach each other, non-adiabatic effects may cause the approximation to break down, necessitating a correction due to the coupling between electronic states. In general, the Born–Oppenheimer framework provides a robust approximation for ground states, but can be less reliable for excited states.

1.1.2 ATOMIC UNITS

Before we delve into the form of the Hamiltonian, we first comment on the concept of atomic units (a.u.), which will be used extensively throughout this thesis, and are implied unless otherwise specified. The advantage of using atomic units lies in their definition, whereby four fundamental constants, namely electronic mass m_e, elementary charge e, the reduced Planck constant $\hbar$, and the electrostatic constant $k_\mathrm{e} = 1/4\pi\varepsilon_0$, are assigned the value of unity.

With this convention, numerous physical constants and derived units can be reduced from mathematically complicated combinations of the above quantities in SI units, to unity or similarly simple forms in atomic units. A summary of units pertinent to this work is given in Table 1.1.

Certain care must be taken in the interpretation of quantities expressed

Table 1.1: Definitions of fundamental and derived atomic units, each with a value of unity, relevant to this work.

Name	Symbol	Approx. value in common units
Electron mass	m_e	$9.109 \times 10^{-31}\,\mathrm{kg}$
Elementary charge	e	$1.602 \times 10^{-19}\,\mathrm{C}$
Reduced Planck's constant	$\hbar = \dfrac{h}{2\pi}$	$1.055 \times 10^{-34}\,\mathrm{J\,s}$
Electrostatic constant	$k_\mathrm{e} = \dfrac{1}{4\pi\varepsilon_0}$	$8.988 \times 10^{9}\,\mathrm{N\,m^2\,C^{-2}}$
Bohr radius; "bohr" [length]	$a_0 = \dfrac{4\pi\varepsilon_0\hbar^2}{m_\mathrm{e}e^2}$	$0.529\,\text{Å}$
Hartree [energy]	$E_\mathrm{h} = \dfrac{m_\mathrm{e}e^4}{(4\pi\varepsilon_0\hbar)^2} = \dfrac{e^2}{a_0}$	$27.211\,\mathrm{eV}$
Electric potential	$\dfrac{E_\mathrm{h}}{e}$	$27.211\,\mathrm{V}$
Electric dipole moment	ea_0	$2.542\,\mathrm{D}$

in atomic units—since they reduce to unity and effectively cancel each other out they are often thought of and expressed as unitless, or otherwise under the catch-all notation of "a.u.". This can lead to a certain dimensional ambiguity when mathematical operations are performed; it is important to note that the same rules of dimensional analysis must apply as with traditional units, in particular the homogeneity of units.

1.1.3 Electronic Schrödinger equation

Within the Born–Oppenheimer approximation, we can reduce equation (1.2) to an *electronic* time-independent Schrödinger equation, which we formally restate for an N-electron system as

$$\hat{H}_\mathrm{e}\psi_\mathrm{e}(\boldsymbol{x}^N;\boldsymbol{R}) = E_\mathrm{e}(\boldsymbol{R})\psi_\mathrm{e}(\boldsymbol{x}^N;\boldsymbol{R})\,. \tag{1.3}$$

The "e" subscript refers to the *electronic* quantities, $\boldsymbol{x}^N = \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ refers to the dependence on each of the electronic space–spin coordinates, and $\boldsymbol{R}$ refers to the parametric dependence on the (fixed) positions of the nuclei. For the remainder of the work we will be working exclusively with this electronic problem, and so we simplify the notation back to

$$\hat{H}\psi = E\psi\,. \tag{1.4}$$

The Hamiltonian $\hat{H}$ is given, in atomic units, by

$$\hat{H} = -\frac{1}{2}\sum_{i}^{N}\nabla_i^2 + \sum_{i}^{N} v(\boldsymbol{r}_i) + \sum_{i<j}^{N}\frac{1}{r_{ij}}\,, \tag{1.5}$$

where

$$v(\boldsymbol{r}_i) = -\sum_{\alpha}\frac{Z_\alpha}{r_{i\alpha}} \tag{1.6}$$

is the external potential acting on electron i due to nuclear charges Z_α. For brevity, we define

$$\hat{H} = \hat{T} + \hat{V}_{\text{en}} + \hat{V}_{\text{ee}} \tag{1.7}$$

where $\hat{T}$, $\hat{V}_{\text{en}}$, and $\hat{V}_{\text{ee}}$ correspond to the three terms in equation (1.5), and are identified as the kinetic, electron–nucleus attraction, and electron–electron repulsion operators, respectively. Note that the potential in equation (1.6) will contain additional terms in the presence of external fields.

The total energy is then given by the sum of the electronic energy E and the nucleus–nucleus repulsion energy V_{nn},

$$V_{\text{nn}} = \sum_{\alpha<\beta}\frac{Z_\alpha Z_\beta}{R_{\alpha\beta}}\,. \tag{1.8}$$

The result is independent of whether this quantity is added after solving equation (1.4) for E, or whether it is included in the definition of $\hat{H}$ itself and the total energy determined directly.

Many acceptable solutions exist to equation (1.4), comprising the eigenfunctions ψ with their corresponding eigenvalues E; the ground state wavefunction and energy are denoted ψ_0 and E_0, respectively. The eigenfunctions ψ form a complete, orthonormal basis, in terms of which any observable state ψ'—which may or may not be an eigenfunction of $\hat{H}$—may be expanded. It follows that the expectation value (the mean of many measurements) of the energy, given by

$$E[\psi'] = \langle\hat{H}\rangle = \frac{\langle\psi'|\hat{H}|\psi'\rangle}{\langle\psi'|\psi'\rangle}\,, \tag{1.9}$$

is an upper bound to the exact ground state energy E_0,

$$E[\psi'] \geqslant E_0\,. \tag{1.10}$$

The ground state energy is given only if equation (1.9) is evaluated with the exact ground state wavefunction ψ_0. We can therefore define the ground state

wavefunction and energy as a full minimisation of the functional $E[\psi]$ with respect to all allowed N-electron wavefunctions,

$$E_0 = \min_{\psi} E[\psi]\,. \tag{1.11}$$

1.2 Hartree–Fock theory

The simplest appropriate approximation to the wavefunction comprises a product of N one-electron spin orbitals $\chi_i(\boldsymbol{x})$, antisymmetrised to satisfy the Pauli principle,[17] in the form of a Slater determinant[18,19]

$$\psi_{\text{HF}} = \frac{1}{\sqrt{N!}} \det[\chi_1(\boldsymbol{x}_1), \chi_2(\boldsymbol{x}_2), \ldots, \chi_N(\boldsymbol{x}_N)]\,, \tag{1.12}$$

where each χ_i is the product of a spatial orbital dependent on the coordinate $\boldsymbol{r}$ and a spin function accounting for α and β spin states, and the prefactor ensures the normalisation condition $\langle\psi_{\text{HF}}|\psi_{\text{HF}}\rangle = 1$. Hartree–Fock (HF) theory[20–24] takes this Slater determinant as an approximation for the true wavefunction, by finding the set of orbitals that minimises the expectation value

$$E_{\text{HF}} = \langle\psi_{\text{HF}}|\hat{H}|\psi_{\text{HF}}\rangle \tag{1.13}$$

$$= \sum_i^N H_i + \frac{1}{2}\sum_{i,j}^N (J_{ij} - K_{ij}) \tag{1.14}$$

$$= \sum_i^N \langle i|\hat{h}|i\rangle + \frac{1}{2}\sum_{i,j}^N \langle ij|ij\rangle - \langle ij|ji\rangle\ , \tag{1.15}$$

where we introduce shorthand notation for one- and two-electron integrals $\langle i|j\rangle = \langle\chi_i|\chi_j\rangle$, and $\langle ij|kl\rangle = \langle\chi_i\chi_j|r_{12}^{-1}|\chi_k\chi_l\rangle$, respectively, using the Dirac notation such that

$$\langle\chi_i\chi_j|r_{12}^{-1}|\chi_k\chi_l\rangle = \iint \chi_i^*(\boldsymbol{x}_1)\chi_j^*(\boldsymbol{x}_2) r_{12}^{-1} \chi_k(\boldsymbol{x}_1)\chi_l(\boldsymbol{x}_2)\, d\boldsymbol{x}_1\, d\boldsymbol{x}_2. \tag{1.16}$$

The operator $\hat{h}$ corresponds to the one-electron Hamiltonian,

$$\hat{h} = -\frac{1}{2}\nabla^2 + v\,, \tag{1.17}$$

and J_{ij} and K_{ij} correspond to the *Coulomb* and *exchange* integrals, respectively. The former term is the classical Coulomb repulsion of a charge distribution

with itself, whereas the second term arises due to the antisymmetry condition of the Pauli principle, and is a lowering of the energy due to like-spin electrons avoiding one other. Importantly, $J_{ii} = K_{ii}$, and so any terms pertaining to an electron interacting with itself cancel, correctly, to zero; the consequence of the failure of approximate DFT (Chapter 2) functionals to replicate this is of direct relevance to this thesis.

The minimisation of equation (1.13)—which yields the best approximate wavefunction of this form due to the variational principle—proceeds under the constraint of orthonormal orbitals

$$\langle i|j\rangle = \delta_{ij}\,. \tag{1.18}$$

This gives rise to the Hartree–Fock equations,

$$\hat{F}\chi_i(\boldsymbol{x}_1) = \sum_j^N \varepsilon_{ij}\chi_j(\boldsymbol{x}_1)\,, \tag{1.19}$$

where ε_{ij} are the elements of a matrix of Lagrange multipliers, the Fock operator $\hat{F}$ defines the effective Hamiltonian,

$$\hat{F} = \hat{h} + \hat{j} - \hat{k}\,, \tag{1.20}$$

and $\hat{j}$ and $\hat{k}$ are the Coulomb and exchange operators respectively. These are defined, by their effect on orbital $\chi_i(\boldsymbol{x}_1)$, as

$$\hat{j}\,\chi_i(\boldsymbol{x}_1) = \sum_j^N \left\{ \int \frac{\chi_j^*(\boldsymbol{x}_2)\chi_j(\boldsymbol{x}_2)}{r_{12}}\, d\boldsymbol{x}_2 \right\} \chi_i(\boldsymbol{x}_1)\,, \tag{1.21}$$

and

$$\hat{k}\,\chi_i(\boldsymbol{x}_1) = \sum_j^N \left\{ \int \frac{\chi_j^*(\boldsymbol{x}_2)\chi_i(\boldsymbol{x}_2)}{r_{12}}\, d\boldsymbol{x}_2 \right\} \chi_j(\boldsymbol{x}_1)\,. \tag{1.22}$$

The HF (Slater determinant) wavefunction is invariant to a unitary transformation of the orbitals, as is the Fock operator. Furthermore, the matrix of Lagrange multipliers $\boldsymbol{\varepsilon}$ is Hermitian, and so there must exist a transformation (and corresponding set of transformed orbitals) that diagonalises $\boldsymbol{\varepsilon}$. Thus we can write the Hartree–Fock equations in the *canonical* form,

$$\hat{F}\chi_i(\boldsymbol{x}_1) = \varepsilon_i\chi_i(\boldsymbol{x}_1)\,. \tag{1.23}$$

The solutions to these equations for an N-electron system comprise a set of eigenfunctions—the orbitals $\{\chi_i\}$—with associated eigenvalues—the orbital energies $\{\varepsilon_i\}$. The N orbitals lowest in energy correspond to the occupied orbitals, whereas the remainder (in principle infinite in number) are the unoccupied, or virtual orbitals.

It must be noted that the eigenvalue equation equation (1.23) is not truly a linear problem. In fact, the Fock operator (through the Coulomb and exchange operators) is itself dependent on the solutions $\{\chi_i\}$ to the problem. The resolution to this apparent contradiction comes in the form of the self-consistent-field (SCF) method, whereby the problem is tackled iteratively, starting from an initial guess for the orbitals.

1.2.1 Roothaan–Hall equations

For an N-electron closed-shell system, a *restricted* (RHF) formalism is customarily employed, where $N/2$ pairs are formed of opposite-spin electrons. Each α-spin electron occupies a spin orbital with an identical spatial component to its β-spin partner—or, equivalently, the HF wavefunction comprises $N/2$ *doubly occupied* spatial orbitals. One can then integrate over the spin functions to derive spatial HF equations,

$$\hat{F}\varphi_i(\boldsymbol{r}) = \varepsilon_i\varphi_i(\boldsymbol{r})\,, \tag{1.24}$$

where $\varphi_i(\boldsymbol{r})$ is the spatial part of $\chi_i(\boldsymbol{x})$.

Although numerical solutions to the HF equations equation (1.24) are tractable for atoms due to their spherical symmetry, the situation is much more complex for molecular systems, and for most practical purposes the orbitals are expanded in a basis set of known spatial functions,

$$\varphi_i(\boldsymbol{r}) = \sum_{\nu} c_{\nu i}\vartheta_\nu(\boldsymbol{r})\,. \tag{1.25}$$

If $\{\vartheta_i\}$ were a complete set, the expansion would be exact, however for practical purposes a finite set must be used; the choice of basis set thus has a direct effect on both the quality of the expansion and the computational complexity, and is an active and subjective area of research.

By substituting equation (1.25) into the HF equation equation (1.24), multiplying on the left by ϑ_μ^* and integrating over $\boldsymbol{r}$, we obtain

$$\sum_\nu c_{\nu i} \langle \vartheta_\mu | \hat{F} | \vartheta_\nu \rangle = \varepsilon_i \sum_\nu c_{\nu i} \langle \vartheta_\mu | \vartheta_\nu \rangle \ . \tag{1.26}$$

We define the Fock matrix $\boldsymbol{F}$ and the overlap matrix $\boldsymbol{S}$ as

$$F_{\mu\nu} = \langle \vartheta_\mu | \hat{F} | \vartheta_\nu \rangle \ , \tag{1.27}$$

and

$$S_{\mu\nu} = \langle \vartheta_\mu | \vartheta_\nu \rangle \ , \tag{1.28}$$

where the latter arises as there is no orthonormality constraint on the basis functions (for an orthonormal basis set, $\boldsymbol{S}$ is the identity matrix). The Roothaan–Hall[25,26] equations of equation (1.26) can then be simply represented as

$$\boldsymbol{Fc} = \boldsymbol{Sc\varepsilon} \, , \tag{1.29}$$

a set of matrix equations soluble using conventional techniques.

1.2.2 BASIS SETS

There are two approaches to the choice of basis set that dominate the field. Plane wave basis sets are common in calculations on periodic systems, however for isolated gas-phase molecules—and all the calculations relevant to this work—by far the most prevalent choice is for atom-centered basis functions. The idea stems from the knowledge that the electronic Schrödinger equation can be solved exactly for hydrogenic systems, coupled with the assumption that the one-electron Hartree–Fock orbitals will be similar to these solutions, and that atoms in molecules will behave similarly to lone atoms.

Atomic orbitals (AOs) are chosen to approximate the hydrogenic solutions, centered on each of the nuclei. The first proposed approximations[27,28] were Slater-type orbitals (STOs), which take the form (for a function centred on the origin $\boldsymbol{r}_0$)

$$\vartheta^{\mathrm{STO}}(\boldsymbol{r}) = \mathcal{N} |\boldsymbol{r} - \boldsymbol{r}_0|^{n-1} \, \mathrm{e}^{-\zeta |\boldsymbol{r} - \boldsymbol{r}_0|} \, , \tag{1.30}$$

where $\mathcal{N}$ is a normalisation constant, n is the (effective) principle quantum number, and ζ is an adjustable parameter, the *Slater orbital exponent*. Note

that we have omitted the system-independent angular part of the function, which is universally given by the set of spherical harmonics.

The Slater functions exhibit a form that can lead to highly accurate results, however the evaluation of the four-centre integrals described by equation (1.16) over these orbitals is a difficult computational prospect, and the calculation quickly becomes difficult as system size increases. By far the most common functions in use today are the Gaussian-type orbitals (GTOs) introduced by Boys.[29] In Cartesian form, these are written

$$\vartheta^{\mathrm{GTO}}(\boldsymbol{r}) = \mathcal{N}(x-x_0)^i(y-y_0)^j(z-z_0)^k\,\mathrm{e}^{-\alpha|\boldsymbol{r}-\boldsymbol{r}_0|^2}\,, \tag{1.31}$$

where the sum $i+j+k$ is analogous to the angular momentum quantum number l, and α is the *Gaussian orbital exponent*.

In contrast to the STOs, the Gaussian functions exhibit a zero slope at the origin rather than the cusp present in the true hydrogenic orbitals, and decay much more rapidly at long range. For this reason STOs give a better qualitative description of the hydrogenic orbitals than GTOs; a more accurate description can be built up for the latter by taking a linear combination of many Gaussian primitives, leading to *contracted* Gaussian functions. This is made computationally possible by the greater ease of evaluating integrals over the GTOs: the product of two Gaussian functions at different centres reduces to a single Gaussian function at a centre between the two.

The Gaussian functions can, as written above, be implemented in their Cartesian form and, indeed, this is often simpler. However a subtlety arises for $l \geqslant 2$, where a greater number of Cartesian Gaussian functions arise than the corresponding spherical harmonics. As an example, the "d" orbitals ($l=2$) comprise five spherical harmonics, but six combinations of i, j, and k. Linear combinations of the Cartesian GTOs give five combinations corresponding to the five spherical GTOs, whereas the sixth (superfluous) combination exhibits the angular behaviour of an s-orbital. Spurious functions such as this are somewhat inefficient, and it may be preferable to transform the GTOs into the spherical harmonic basis and so reduce the number of basis functions.

A final consideration to note is the number of functions that are required for the basis set to be sufficiently flexible to produce accurate results. In a so-called *minimal* basis set, there exists a single basis function (typically a contracted Gaussian function) to describe each AO. Improvements can then be

made by including additional functions (both of the same and of higher angular momenta) to increase the flexibility of the basis set. Polarisation functions typically take the form of orbitals with higher angular momenta than those of the occupied AOs—these have the effect of distorting the shape of the AOs, accounting for the distortion of the electron density due to, for example, other atoms in the molecule. Polarisation functions are also vital to the description of electron correlation, in order to correctly describe the electron–electron cusp arising from the reduction in the probability of two electrons occupying the same space.

Additional diffuse functions can also be included for cases where the electron density is more spread out, such as in excited states and anions. These typically take the form of single, shallow GTOs with small exponents and improve the description at larger distances from the nucleus.

1.2.3 OPEN-SHELL SYSTEMS

For open-shell systems two formalisms arise. In *restricted open-shell* Hartree–Fock (ROHF), pairs of opposite-spin electrons exist in doubly occupied "closed-shell" orbitals, whilst the remainder exist in singly occupied "open-shell" orbitals. Conversely in *unrestricted* Hartree–Fock (UHF), the spatial components of orbitals for opposite-spin electrons are allowed to differ. The orbitals are obtained from the Pople–Nesbet equations,[30] which are α and β analogues of the Roothaan–Hall equations.

In principle, the UHF formalism for closed-shell systems should reduce to the RHF case, although treatment of cases such as dissociation into open-shell fragments will be handled differently. For open-shell systems, UHF tends to give a lower variational energy than ROHF due to its better description of, for example, an unpaired $2s\alpha$ electron exhibiting a different interaction with either a (paired) $1s\alpha$ or $1s\beta$ electron. That being said, care must be taken, as whilst the ROHF wavefunction is an eigenfunction of the $\hat{S}^2$ spin operator, the UHF wavefunction is not, and may exhibit varying degrees of spin contamination.

1.3 Correlated methods

Through the variational principle, the HF energy is an upper bound on the exact energy. Aside from the one-electron case for which it is exact, it remains an approximation due to its mean-field treatment of the interaction between an electron with the others in the system, through modelling the wavefunction as a single determinant. In particular, the mean-field approach fails to account for the electron–electron cusp of opposite-spin electrons (although some correlation between like-spin electrons is modelled), and a single determinant cannot accurately describe certain situations such as a molecule approaching dissociation.

The correlation energy is defined as the difference between the exact energy and the HF energy,

$$E_{\mathrm{C}} = E_0 - E_{\mathrm{HF}} \leqslant 0\,. \tag{1.32}$$

Much of the research in the field of electronic structure theory has revolved around calculating and incorporating this correlation energy, in order to solve problems with quantitative accuracy. For the remainder of this chapter, we briefly discuss a number of approaches, before justifying the need for a more computationally accessible alternative, which will be the focus of the remainder of the thesis.

1.3.1 configuration interaction

The justification for the method of configuration interaction arises from the ability[31] to express the exact wavefunction as a linear combination of all possible N-electron Slater determinants that arise from a complete set of spin orbitals,

$$|\psi_{\mathrm{CI}}\rangle = |\psi_{\mathrm{HF}}\rangle + \sum_{a,i} c_i^a\,|\psi_i^a\rangle + \sum_{a<b,i<j} c_{ij}^{ab}\,|\psi_{ij}^{ab}\rangle + \ldots \tag{1.33}$$

$$= |\psi_{\mathrm{HF}}\rangle + \sum_{i,a} c_i^a \hat{\tau}_i^a\,|\psi_{\mathrm{HF}}\rangle + \frac{1}{4}\sum_{a,b,i,j} c_{ij}^{ab}\hat{\tau}_{ij}^{ab}\,|\psi_{\mathrm{HF}}\rangle + \ldots \tag{1.34}$$

$$= (1 + \hat{C}_1 + \hat{C}_2 + \ldots)\,|\psi_{\mathrm{HF}}\rangle\,. \tag{1.35}$$

The $\hat{C}_1$ (singles) excitation operator generates all possible singly excited determinants by changing the orbital occupancy of the wavefunction, through the operator $\hat{\tau}_i^a$, which "moves" an electron in the ith occupied orbital to the ath

virtual orbital; the coefficients c_i^a (amplitudes) are determined by minimising the electronic energy. The $\hat{C}_2$ (doubles) operator is analogous for doubly excited determinants, with the series increasing until all possible Slater determinants are included, the so-called full configuration interaction (FCI) limit.

FCI is exact (within a given basis set—approximations are still introduced by the need for a finite basis set), but the vast number of terms that emerge as the number of electrons increases means it is impossible to compute for all but the smallest of systems. In practice, one must truncate the expansion to a finite order. Inclusion of singles only (CIS) is no different to Hartree–Fock itself: the singly excited determinants do not couple with the ground state, as a consequence of Brillouin's theorem.[32] As a result, to see any improvement one must at least include singles and doubles (CISD), and this remains the most commonly used truncation.

One disadvantage of truncated CI is that it is not size-extensive. Consider two non-interacting He atoms: for a method to be size-extensive, a calculation on the two-atom, four-electron system should give an identical energy to twice that of a single isolated He atom. The CISD method is equivalent to FCI for a two-electron He atom: single and double excitations generate all possible determinants and so the method is exact; doubling the energy gives the exact energy for the two non-interacting atoms. A CISD calculation on the four-electron system with two He atoms, however, neglects the determinants formed by triple and quadruple excitations and so does not give the same, exact, energy. Although the size-extensivity error can be reduced,[33] its presence restricts the usefulness of truncated CI, leading to the consideration of the size-extensive alternative of coupled-cluster theory.

1.3.2 COUPLED-CLUSTER THEORY

Whereas CI represents the wavefunction as a sum of excitation operations

$$|\psi_{\text{CI}}\rangle = (1 + \sum_{\mu} c_\mu \hat{\tau}_\mu)\,|\psi_{\text{HF}}\rangle\ , \tag{1.36}$$

the coupled-cluster (CC) method instead expands the wavefunction as a product,[34–36]

$$|\psi_{\text{CC}}\rangle = \prod_{\mu}(1 + t_\mu \hat{\tau}_\mu)\,|\psi_{\text{HF}}\rangle \tag{1.37}$$

$$= \mathrm{e}^{\hat{T}}\,|\psi_{\text{HF}}\rangle \tag{1.38}$$

$$= (1 + \hat{T} + \frac{1}{2!}\hat{T}^2 + \ldots)\,|\psi_{\text{HF}}\rangle\ , \tag{1.39}$$

which holds due to the property $\hat{\tau}_\mu^2 = 0$. The excitation operator is given by

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \ldots\ , \tag{1.40}$$

where

$$\hat{T}_1 = \sum_{a,i} t_i^a \hat{\tau}_i^a\ , \tag{1.41}$$

$$\hat{T}_2 = \sum_{a<b,i<j} t_{ij}^{ab} \hat{\tau}_{ij}^{ab}\ , \tag{1.42}$$

and so on. As with FCI, a full expansion of all possible excited determinants gives the exact wavefunction for the given basis set, however it is when the expansion is truncated that the advantages of CC manifest.

Consider the prevalent truncation including only the single and double excitations, CCSD, where $\hat{T} = \hat{T}_1 + \hat{T}_2$. The inclusion of singly excited determinants is similar to CI, with each contribution possessing an associated single-excitation amplitude t_a^i. Doubly excited determinants, however, can arise in two ways: both from the *connected* contribution of the $\hat{T}_2$ operator, with associated double-excitation amplitudes t_{ab}^{ij}, and from the *disconnected* contribution of the $\hat{T}_1$ operator acting twice. This disconnected contribution is instead associated with a product of single-excitation amplitudes $t_i^a t_j^b$. In a similar manner, disconnected contributions appear from triple, quadruple, and higher order excitations.

Whilst these disconnected contributions do not give an exact treatment of the higher-order excitations, their inclusion is beneficial to the overall accuracy of the method. Importantly, returning to the example of two non-interacting He atoms, the presence of the disconnected contributions accounts for all possible excitations, and so the method is size-extensive.

The CCSD method is a popular benchmarking tool, and can presently be applied to molecules of reasonable size. Systematic improvement, however, in order to achieve true quantitative accuracy by inclusion of triples (CCSDT[37]) and higher terms, is again hindered by an unmanageable computational cost for all but the smallest of systems. Approximations to both CCSD and CCSDT have been developed in the form of CC2[38,39] and CC3,[40,41] which mitigate some of the cost of their parent methods.

A variant of CCSD is formulated in terms of the Brueckner orbitals[42,43] rather than the HF orbitals, and is termed the *Brueckner Doubles* (BD)[44] method. The Brueckner orbitals are linear combinations of the HF orbitals that give rise to single-excitation amplitudes of zero in the coupled-cluster formulation. The benefit of this approach is in mitigating some of the complexity of CCSD, along with potential problems associated with large single-excitation amplitudes, without sacrificing its accuracy.

A final important note is that the CC methods are not variational, and may overestimate the correlation energy. The typical error, however, is much lower than that caused by the non-size-extensivity of truncated CI.

1.3.3 PERTURBATION THEORY

An alternative approach, which—like coupled cluster—is (in principle) systematic, size-extensive at any level of truncation, but not variational, can be identified by considering perturbation theory. We represent the true Hamiltonian for a system as $\hat{H} = \hat{H}_0 + \lambda\hat{V}$, where $\hat{H}_0$ is a reference Hamiltonian whose eigenvalues and eigenfunctions are known, and $\hat{V}$ is a perturbation that transforms the reference system into the true system through the parameter λ ($\lambda = 0$ indicates the zeroth order, unperturbed system whereas $\lambda = 1$ describes the true system).

If $\hat{H}_0$ is well chosen then the perturbation will be small, and we can approximate the exact wavefunction and energy by expanding as a Taylor series in λ. Although we present only the main results relevant to this work, a full treatment is again given in Refs 2 and 3.

Møller–Plesset[45] perturbation theory (MPPT) chooses $\hat{H}_0$ to be a sum over the one-electron Fock operators of equation (1.20), with the perturbation given by $\hat{V} = \hat{H} - \hat{H}_0 = \hat{H} - \sum_i \hat{F}(r_i)$. The zeroth order energy is simply the

expectation value of $\hat{H}_0$,

$$E^{(0)} = \langle \psi_{\text{HF}} | \hat{H}_0 | \psi_{\text{HF}} \rangle \ , \tag{1.43}$$

where ψ_{HF} is the ground-state HF wavefunction, whilst the first order correction is given by

$$E^{(1)} = \langle \psi_{\text{HF}} | \hat{V} | \psi_{\text{HF}} \rangle \ . \tag{1.44}$$

It is immediately clear that the Hartree–Fock energy is equivalent to the sum of these two terms, *i.e.* the first-order truncation of MPPT,

$$E_{\text{HF}} = \langle \psi_{\text{HF}} | \hat{H} | \psi_{\text{HF}} \rangle = \langle \psi_{\text{HF}} | \hat{H}_0 + \hat{V} | \psi_{\text{HF}} \rangle = E^{(0)} + E^{(1)} \, . \tag{1.45}$$

It can be shown that the second order correction to the energy is given by

$$E^{(2)} = \frac{1}{4} \sum_{a,b,i,j} \frac{| \langle ij|ab\rangle - \langle ij|ba\rangle |^2}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \tag{1.46}$$

and truncating here yields the second-order MPPT approximation, denoted MP2. Although MP2 is a commonly used method, further terms can be systematically added by truncating at higher orders, however doing so does not, in general, produce a systematic improvement worthy of the increasing computational complexity.

The principles of MPPT can also be applied to the coupled-cluster methods of Section 1.3.2. Applying the MP-type perturbation to the CCSD wavefunction, such that the triples correction is introduced perturbatively[46] yields the CCSD(T) method. This useful approach provides much of the improved accuracy of the full CCSDT method, without a large part of the significant increase to computational cost. A similar perturbative approach can incorporate triples into the BD method, to give a method denoted BD(T).

1.3.4 Basis Set Dependence

A final comment must be made on the dependence of these correlated wavefunction methods on the choice of basis set. In particular, a great many polarisation functions are required to accurately model the electron–electron cusp, resulting in very large basis sets. This, in part, stems from the need to describe not only the ground-state (HF) determinant, but also excited-state determinants,

which involve occupation of orbitals that are virtual in the ground state. The accurate description of a wavefunction incorporating an increasing number of excited-state determinants requires an ever-increasing flexibility in the basis set, and the convergence towards the complete basis limit is slow. Similarly, any small increase in system size necessitates a large increase in basis set flexibility.

In general, correlated methods scale poorly with system size—CCSD(T), for example, formally scales as M^7, where M is a measure of the system size related to the number of basis functions. Since the number of basis functions required for an accurate calculation itself increases with system size, calculations can rapidly become intractable, and the use of these methods—using presently available computational techniques—is unfeasible for all but the smallest of systems.

As an illustration, the limit to date of full CI has been systems of no more than around ten electrons, standard CCSD(T) implementations would take weeks or months to model several tens of electrons, whereas more efficient algorithms have permitted calculations on up to several hundred electrons. Despite rapid increases in computational power and memory, along with many technical advances in the optimisation of methods and algorithms, application of these correlated methods to systems of several thousand electrons or more are likely to remain inaccessible for the foreseeable future, and it is this that leads us to consider methods formulated around a much simpler, more accessible quantity: the electron density.

2

Density functional theory

Density functional theory is introduced as a formally exact alternative to correlated *ab initio* electronic structure methods. The electron density is shown, through the Hohenberg–Kohn theorems, to contain sufficient information to describe and determine all properties of the system. The Kohn–Sham method is then introduced as a practical route to solving electronic structure problems, by approximating the exchange–correlation functional. Finally, methods of approximating this functional are discussed, along with ongoing challenges in their development.

Chapter 1 introduced approaches to approximately solve the Schrödinger equation by approximating the exact ground-state wavefunction ψ_0. Whilst such methods can, in principle, produce very accurate results, in practice it becomes very difficult to perform calculations on even modest systems, owing to the huge complexity of the N-electron wavefunction $\psi(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)$. This complexity stems from its dependence on the four (three spatial and one spin) coordinates of each electron in the system along with, implicitly, the positions of the nuclei. This, coupled with the increasingly large basis sets needed to correctly model the electron–electron cusp, led many to investigate whether a simpler quantity could be used to determine the required information.

One such quantity was the one-electron density $\rho(\boldsymbol{r})$, and early models involving ρ date back to the late 1920s, with the work of Thomas, Fermi and Dirac.[47–49] At the time, there was no formal proof that an exact theory could

be derived in terms of ρ, and these models did not gain wide acceptance until much later, when in 1964 Hohenberg and Kohn provided such a proof. Despite this, their decision to use the electron density ρ—more properly the electron *probability* density—was arguably an intuitive one, as reputedly observed by E. Bright Wilson following a presentation of Hohenberg and Kohn's work (the latter is discussed in Section 2.2). He noted that cusps in the density indicate the positions of the nuclei, and that the slope of the cusp is related to the nuclear charge by

$$\frac{\partial}{\partial r_A}\bar{\rho}(r_A)|_{r_A=0} = -2Z_A\bar{\rho}(0)\,, \tag{2.1}$$

for a given nucleus A, where $\bar{\rho}$ is the spherical average of the density. This, coupled with the knowledge that the density integrates to the total number of electrons

$$\int \rho(\boldsymbol{r})\,d\boldsymbol{r} = N\,, \tag{2.2}$$

suggests that all the information required to determine the Hamiltonian, and thus all the desired properties of a system, is contained within the density.

2.1 Thomas–Fermi theory

Conceptually, the idea was simple: to express each of the terms defining the total energy associated with the Hamiltonian operator in equation (1.7) as a *functional* of the density, rather than the wavefunction, such that

$$E[\rho] = T[\rho] + V_{\text{en}}[\rho] + V_{\text{ee}}[\rho] \tag{2.3}$$

$$= T[\rho] + \int \rho(\boldsymbol{r})v(\boldsymbol{r})\,d\boldsymbol{r} + V_{\text{ee}}[\rho]\,, \tag{2.4}$$

where $T[\rho]$ is the electronic kinetic energy, $V_{\text{en}}[\rho]$ is the electron–nuclear interaction energy, and $V_{\text{ee}}[\rho]$ is the electron–electron repulsion energy. Note that, simplistically, a functional $F[f]$ (denoted with square brackets), differs from a function $f(x)$, in that the latter takes a number as input and returns a number, whereas the former takes a *function* as input and returns a number.

Using statistical arguments to model the distribution of electrons in an atom, Thomas[47] and Fermi[48] derived an expression for the total kinetic energy of a *non-interacting* system in terms of the electron density,

$$T_{\text{TF}}[\rho] = C_{\text{T}}\int \rho^{5/3}(\boldsymbol{r})\,d\boldsymbol{r}, \quad \text{where} \quad C_{\text{T}} = \frac{3}{10}(3\pi^2)^{2/3}\,. \tag{2.5}$$

This is an example of the modern idea of a *local density approximation*, where conditions appropriate for a uniform electron gas (UEG) are applied to local systems.

The original Thomas–Fermi model approximated V_{ee} using only the classical Coulomb potential energy,

$$J[\rho] = \frac{1}{2} \iint \frac{\rho(\boldsymbol{r}_1)\rho(\boldsymbol{r}_2)}{r_{12}} \, d\boldsymbol{r}_1 \, d\boldsymbol{r}_2 \, . \tag{2.6}$$

Of course, this neglects completely the non-classical part, and the Thomas–Fermi–Dirac model[49] improves the approximation by adding to V_{ee} the exchange energy of the UEG,

$$E_{\text{x}}[\rho] = C_{\text{x}} \int \rho^{4/3}(\boldsymbol{r}) \, d\boldsymbol{r}, \quad \text{where } C_{\text{x}} = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} . \tag{2.7}$$

The true electron density of an atom or molecule does not resemble that of the UEG, and so the approximations remain somewhat over-simplified. It is likely for this reason—along with the failure to match the accuracy of existing approximations—that the model was initially seen as unimportant for any quantitative application. The kinetic energy term, in particular, was problematic. From the Virial theorem, the kinetic energy for an atom is the negative of the total energy, and is thus relatively large. As a result, even small errors in its approximation can lead to large errors in the result obtained.

In the context of Hohenberg and Kohn's proof, however, the perspective changed dramatically. Now, the TFD model was an approximation to an exact theory, offering important insights into the potential for improved approximations. Indeed, the principles of modelling a non-interacting system, and of a local density approximation, became vital components of Kohn–Sham theory and early approximations of the exchange–correlation energy (Section 2.4).

2.2 The Hohenberg–Kohn theorems

As can be seen from equation (1.5), the Hamiltonian for a system is completely defined by the number of electrons N and the external potential $v(\boldsymbol{r})$, and so with the knowledge of these quantities one can, in principle, determine all properties of the ground state. The legitimacy of replacing N and $v(\boldsymbol{r})$ with

$\rho(\boldsymbol{r})$—as with the TFD explicit functionals of the density—was proven in 1964 by Hohenberg and Kohn.[50]

2.2.1 First Hohenberg–Kohn theorem

Their deceptively simple *reductio ad absurdum* proof shows that two external potentials differing by more than an additive constant cannot be associated with the same density, or in other words $\rho(\boldsymbol{r})$ uniquely determines $v(\boldsymbol{r})$. Since N is also determined trivially by quadrature, all terms in the electronic Hamiltonian can be written as functionals of the density.

Consider two external potentials, $v(\boldsymbol{r})$ and $v'(\boldsymbol{r})$, differing by more than a constant, which give the same ground-state density $\rho(\boldsymbol{r})$. These, in turn, give rise to two Hamiltonians, $\hat{H}$ and $\hat{H}'$, with the same density but different normalised wavefunctions ψ and ψ', and corresponding ground-state energies E and E'. Taking ψ' as a trial wavefunction for $\hat{H}$ gives, from the variational principle,

$$E_0 < \langle\psi'|\hat{H}|\psi'\rangle = \langle\psi'|\hat{H}'|\psi'\rangle + \langle\psi'|\hat{H} - \hat{H}'|\psi'\rangle \tag{2.8}$$

$$= E_0' + \int \rho(\boldsymbol{r})\big[v(\boldsymbol{r}) - v'(\boldsymbol{r})\big]\,d\boldsymbol{r}\,. \tag{2.9}$$

Similarly, taking ψ as a trial wavefunction for $\hat{H}'$ gives

$$E_0' < \langle\psi|\hat{H}'|\psi\rangle = \langle\psi|\hat{H}|\psi\rangle + \langle\psi|\hat{H}' - \hat{H}|\psi\rangle \tag{2.10}$$

$$= E_0 - \int \rho(\boldsymbol{r})\big[v(\boldsymbol{r}) - v'(\boldsymbol{r})\big]\,d\boldsymbol{r}\,. \tag{2.11}$$

Summing these two inequalities results in $E_0 + E_0' < E_0' + E_0$, which is a clear contradiction. Thus the original postulate—that two differing potentials can be associated with the same density—must be false.

In other words, $v(\boldsymbol{r})$ is uniquely determined by $\rho(\boldsymbol{r})$ (to within a trivial additive constant). Since N is determined by quadrature,

$$N = \int \rho(\boldsymbol{r})\,d\boldsymbol{r}\,, \tag{2.12}$$

it must follow that $\rho(\boldsymbol{r})$ is able to determine all properties of the ground state. We can thus write the total electronic energy as a functional of the density,

$$E[\rho] = \int \rho(\boldsymbol{r})v(\boldsymbol{r})\,d\boldsymbol{r} + F_{\text{HK}}[\rho]\,, \tag{2.13}$$

where

$$F_{\text{HK}}[\rho] = T[\rho] + V_{\text{ee}}[\rho] = \langle\psi|\hat{T} + \hat{V}_{\text{ee}}|\psi\rangle \ . \tag{2.14}$$

It may also be noted that $V_{\text{ee}}[\rho]$ can be written as the classical Coulomb repulsion (in terms of the density) $J[\rho]$, added to a non-classical term. This latter term is very important, and will be a major component of the exchange–correlation energy defined and discussed below.

2.2.2 SECOND HOHENBERG–KOHN THEOREM

The second theorem of Hohenberg and Kohn[50] introduces the energy variational principle in terms of the density. Consider a trial density $\tilde{\rho}(\boldsymbol{r}) \geqslant 0$, where $\int \tilde{\rho}(\boldsymbol{r})\, d\boldsymbol{r} = N$. From the first Hohenberg–Kohn theorem, there is a unique potential $\tilde{v}(\boldsymbol{r})$, and hence wavefunction $\tilde{\psi}$, associated with $\tilde{\rho}(\boldsymbol{r})$. This wavefunction can in turn be taken as a trial function for the system of interest—with external potential $v(\boldsymbol{r})$—such that

$$\langle\tilde{\psi}|\hat{H}|\tilde{\psi}\rangle = \int \tilde{\rho}(\boldsymbol{r}) v(\boldsymbol{r})\, d\boldsymbol{r} + F_{\text{HK}}[\tilde{\rho}] = E[\tilde{\rho}] \geqslant E[\rho]\,. \tag{2.15}$$

Thus, for a trial density $\tilde{\rho}(\boldsymbol{r})$,

$$E_0 \leqslant E[\tilde{\rho}]\,, \tag{2.16}$$

where E_0 is the ground-state energy.

To obtain the ground-state energy, therefore, the energy must be minimised with respect to variations in the density, subject to the constraint that the number of electrons remains constant, equation (2.12). This leads to the stationary condition

$$\frac{\delta}{\delta\rho(\boldsymbol{r})}\left(E[\rho] - \mu\left[\int \rho(\boldsymbol{r})\, d\boldsymbol{r} - N\right]\right) = 0\,, \tag{2.17}$$

or

$$\frac{\delta E[\rho]}{\delta\rho(\boldsymbol{r})} - \mu = 0\,, \tag{2.18}$$

where $\delta/\delta\rho(\boldsymbol{r})$ denotes a functional derivative, *i.e.* the change in the functional in response to a local change in the input function. The Lagrange multiplier μ is characteristic of the system and is termed the (electronic) *chemical potential*, measuring the escaping tendency of an electron from the equilibrium

system.[51,52] Through its link with electronegativity,[51–53] this quantity has particular significance in conceptual DFT, which will be discussed in Chapter 4.

Combining equations (2.13) and (2.18) gives the Euler–Lagrange equation

$$\mu = \frac{\delta E[\rho]}{\delta \rho(\boldsymbol{r})} \tag{2.19}$$

$$= v(\boldsymbol{r}) + \frac{\delta F_{\text{HK}}[\rho]}{\delta \rho(\boldsymbol{r})} \tag{2.20}$$

We therefore have a formally exact theory: given the exact $F_{\text{HK}}[\rho]$ we can, in principle, solve equation (2.20) for any given system. Unfortunately, the explicit form of $F_{\text{HK}}[\rho]$ is unknown, and it is around this problem that much of the field of research in DFT has developed in recent years. We have seen already that this term can be approximated—indeed, the early work of Thomas, Fermi and Dirac can be seen as approximations to its components $T[\rho]$ and $V_{\text{ee}}[\rho]$, but even small errors in these approximations can render the theory unusable.

A practical workaround to this problem was provided shortly after, by Kohn and Sham, and this is addressed below. First, though, we discuss another potential problem: that of the trial density.

2.2.3 v- and N-representability of the electron density

The Hohenberg–Kohn theorems were an important breakthrough, demonstrating that the ground-state electron density uniquely determines the properties of the ground state, however there is a subtle setback. We now introduce the concept of a v-representable density—which is defined as a density associated with the antisymmetric wavefunction of a Hamiltonian of the form equation (1.5), with some external potential $v(\boldsymbol{r})$. By necessity, the densities considered in the proofs above are v-representable, but in general a density may not be. This is an important distinction, and so we restate the first Hohenberg–Kohn proof as the assertion that a *v-representable* density has a unique mapping to a single ground-state wavefunction, and thus determines the properties of its ground state. Importantly, the idea that all the ground-state properties can be described as functionals of the electron density only holds—assuming the above proofs are followed—if the density is v-representable.

We know the exact ground-state density for a system will be v-representable, but the variational principle of equation (2.15) breaks down if the trial density is not v-representable. This presents a problem, as many physically reasonable trial densities can, in fact, be non-v-representable.[54,55] It turns out that we can, in fact, reformulate the theory such that the density need only satisfy a weaker condition: that of *N-representability*.

An N-representable electron density is one that can be derived from an anti-symmetric N-electron wavefunction: a necessary condition for v-representability, but weaker than the latter. In other words, v-representable densities form a subset of N-representable densities, and so the exact ground-state density is both v- and N-representable. The conditions for N-representability are known and, indeed, are satisfied for any reasonable density. Specifically, the density must be everywhere positive, integrate to the number of electrons, and be differentiable, *i.e.* $\rho(\boldsymbol{r}) \geqslant 0$, $\int \rho(\boldsymbol{r})\, d\boldsymbol{r} = N$, and $\int |\nabla \rho(\boldsymbol{r})^{1/2}|^2\, d\boldsymbol{r} < \infty$.

2.3 The Levy constrained search approach

In order to avoid the v-representability problem, Levy introduced the constrained search approach,[55,56] which minimises over N-representable densities. First, we redefine the $F_{\text{HK}}[\rho]$ functional in terms of a constrained search approach. Consider two wavefunctions: the true ground state ψ_0, and a second wavefunction ψ', both of which integrate to the ground-state density $\rho_0(\boldsymbol{r})$. From the variational principle,

$$\langle \psi' | \hat{H} | \psi' \rangle \geqslant \langle \psi_0 | \hat{H} | \psi_0 \rangle = E_0 \,, \tag{2.21}$$

where $\hat{H} = \hat{T} + \hat{V}_{\text{ee}} + \hat{V}_{\text{en}}$. Since the contribution of $\hat{V}_{\text{en}}$—due to the external potential—is the same for both wavefunctions, we can write

$$\langle \psi' | \hat{T} + \hat{V}_{\text{ee}} | \psi' \rangle \geqslant \langle \psi_0 | \hat{T} + \hat{V}_{\text{ee}} | \psi_0 \rangle \tag{2.22}$$

and so for all wavefunctions giving the ground-state density $\rho_0(\boldsymbol{r})$, the ground-state ψ_0 is that which minimises the expectation value $\langle \hat{T} + \hat{V}_{\text{ee}} \rangle$. The right-hand side of equation (2.22) is equivalent to the functional $F_{\text{HK}}[\rho_0]$ and so we can define

$$F_{\text{HK}}[\rho_0] = \min_{\psi \to \rho_0} \langle \psi | \hat{T} + \hat{V}_{\text{ee}} | \psi \rangle \ . \tag{2.23}$$

This *constrained-search* definition, so-called because it searches over only a subset of possible wavefunctions: those which give the ground-state density $\rho_0(\boldsymbol{r})$, has been derived without reference to the fact that $\rho_0(\boldsymbol{r})$ is v-representable.

By extending the space of trial wavefunctions, Levy showed that one can arrive at the ground state energy with only the N-representability constraint. Consider a general functional

$$F[\rho] = \min_{\psi\to\rho} \langle\psi|\hat{T} + \hat{V}_{\text{ee}}|\psi\rangle \ , \tag{2.24}$$

where the minimisation is now over any N-representable ρ. By definition, for the ground-state density (which we know to be v-representable), $F_{\text{HK}}[\rho_0] = F[\rho_0]$.

Levy's approach partitions the energy-minimisation procedure into two steps. Firstly the energy is minimised over all wavefunctions ψ which give rise to an N-representable density ρ. Secondly, it is minimised over all possible N-representable densities to determine E_0:

$$\begin{aligned} E_0 &= \min_{\psi} \langle\psi|\hat{T} + \hat{V}_{\text{ee}} + \hat{V}_{\text{en}}|\psi\rangle & (2.25)\\ &= \min_{\rho} \left(\min_{\psi\to\rho} \langle\psi|\hat{T} + \hat{V}_{\text{ee}} + \hat{V}_{\text{en}}|\psi\rangle \right) & (2.26)\\ &= \min_{\rho} \left(\min_{\psi\to\rho} \langle\psi|\hat{T} + \hat{V}_{\text{ee}}|\psi\rangle + \int \rho(\boldsymbol{r}) v(\boldsymbol{r})\, d\boldsymbol{r} \right) & (2.27)\\ &= \min_{\rho} \left(F[\rho] + \int v(\boldsymbol{r})\rho(\boldsymbol{r})\, d\boldsymbol{r} \right) . & (2.28) \end{aligned}$$

The existence of a universal functional $F[\rho]$, for any N-representable density, is thus proven. This functional, combined with the variational principle, therefore provides a rigorous, formally exact density functional theory for determining the ground-state density and energy of a given electronic system.

A problem remains, however, in the difficulty of approximating $F[\rho]$. We still do not know how to formulate the components $T[\rho]$ and $V_{\text{ee}}[\rho]$ in terms of the density, and even small errors can prove catastrophic. Kohn and Sham introduced a working solution to the problem, which remains the most commonly applied approach to practical DFT calculations.

2.4 Kohn–Sham theory

As is now well established, the barrier to converting the formally exact proof that the density can be used in place of the many-electron wavefunction, into a practical scheme for accurately calculating the properties of an arbitrary system, lies in the inability to approximate $F[\rho]$ to the required accuracy. Although the full form of $F[\rho]$ is unknown, the density-dependence of some parts of the energy *are* known. The beauty of the Kohn–Sham approach, presented in 1965,[57] is that it separates the large, unknown, $F[\rho]$ into a relatively large component which we know how to calculate exactly, and a much smaller unknown component for which approximation is still necessary—that way, any remaining errors in the approximation will have a much smaller effect on the total energy.

The key lies in a simple repartitioning of the components of the energy expression, by considering a fictitious system of non-interacting electrons with density ρ. Specifically, the kinetic energy T can be represented as the sum of the kinetic energy T_s of the non-interacting system and a corrective term due to the interactions,

$$T[\rho] = T_\mathrm{s}[\rho] + (\text{interacting term})\,. \tag{2.29}$$

Similarly, a large portion of the electron–electron repulsion energy V_ee is given by the classical Coulomb repulsion of the density,

$$J[\rho] = \frac{1}{2}\iint \frac{\rho(\boldsymbol{r}_1)\rho(\boldsymbol{r}_2)}{r_{12}}\,d\boldsymbol{r}_1\,d\boldsymbol{r}_2\,, \tag{2.30}$$

again leaving a small term to be approximated,

$$V_\mathrm{ee}[\rho] = J[\rho] + (\text{non-classical term})\,. \tag{2.31}$$

We collect the "left-over" terms into a new quantity termed the *exchange–correlation (XC) energy functional*, defined as

$$E_\mathrm{XC}[\rho] = T[\rho] - T_\mathrm{s}[\rho] + V_\mathrm{ee}[\rho] - J[\rho]\,. \tag{2.32}$$

In other words, it consists of the difference between the kinetic energy of the full, interacting system and that of the hypothetical non-interacting system,

added to the difference between the electron–electron interaction energy of the real system and the classical Coulomb energy.

We can therefore rewrite the energy as

$$E[\rho] = \int \rho(\boldsymbol{r}) v(\boldsymbol{r})\, d\boldsymbol{r} + T_{\mathrm{s}}[\rho] + J[\rho] + E_{\mathrm{xc}}[\rho]\,. \tag{2.33}$$

Minimising this energy, as before, with respect to changes in the density (subject to constant N), gives the Euler equation

$$\mu = v(\boldsymbol{r}) + \frac{\delta T_{\mathrm{s}}[\rho]}{\delta\rho(\boldsymbol{r})} + \frac{\delta J[\rho]}{\delta\rho(\boldsymbol{r})} + \frac{\delta E_{\mathrm{xc}}[\rho]}{\delta\rho(\boldsymbol{r})}\,. \tag{2.34}$$

If we define an *effective* potential

$$v_{\mathrm{eff}}(\boldsymbol{r}) = v(\boldsymbol{r}) + \frac{\delta J[\rho]}{\delta\rho(\boldsymbol{r})} + \frac{\delta E_{\mathrm{xc}}[\rho]}{\delta\rho(\boldsymbol{r})}\,, \tag{2.35}$$

then we may write

$$\mu = v_{\mathrm{eff}}(\boldsymbol{r}) + \frac{\delta T_{\mathrm{s}}[\rho]}{\delta\rho(\boldsymbol{r})}\,. \tag{2.36}$$

Importantly equation (2.36), which yields the exact density of the real, fully interacting system, is entirely equivalent to the Euler equation of equation (2.20), but for a system of *non-interacting* electrons moving in an external potential $v_{\mathrm{eff}}(\boldsymbol{r})$ (*i.e.* a system where $T = T_{\mathrm{s}}$ and $V_{\mathrm{ee}} = 0$). In other words, we can determine the density of the real, interacting system by performing a calculation on a non-interacting system, with potential $v_{\mathrm{eff}}(\boldsymbol{r})$.

This is trivial: the Hamiltonian for a non-interacting system is

$$\hat{H} = -\frac{1}{2}\sum_i \nabla_i^2 + \sum_i v_{\mathrm{eff}}(\boldsymbol{r}_i)\,, \tag{2.37}$$

the wavefunction ψ is a Slater determinant comprising the one-electron orbitals $\{\varphi_i\}$, and we can thus determine the orbitals that are the solutions to

$$\left(-\frac{1}{2}\nabla^2 + v_{\mathrm{eff}}(\boldsymbol{r})\right)\varphi_i(\boldsymbol{r}) = \epsilon_i\varphi_i(\boldsymbol{r})\,. \tag{2.38}$$

Returning to the energy expression of equation (2.33), the kinetic energy of the non-interacting system is given exactly by

$$T_{\mathrm{s}}[\rho] = \sum_i \langle\varphi_i| -\frac{1}{2}\nabla^2|\varphi_i\rangle\,, \tag{2.39}$$

and the density by

$$\rho(\boldsymbol{r}) = \sum_{i}^{n} |\varphi_i(\boldsymbol{r})|^2 \,. \tag{2.40}$$

Crucially, the density of the real, interacting system is by definition the same as that of the fictitious system.

This indirect procedure of reintroducing orbitals to solve the problem, as an (equivalent) alternative to finding a direct solution to the Euler equation, is known as Kohn–Sham (KS) DFT. By repartitioning the energy expression into equation (2.33) we are able to calculate all the terms exactly, except for the relatively small E_{XC} component, for which an approximation is still necessary. Whilst some of the elegant simplicity of the theory is lost as a result, it overcomes many of the challenges associated with the direct approach, and as such remains the most popular route for practical DFT calculations.

The similarities to Hartree–Fock theory are immediately apparent, and, indeed, KS-DFT requires a similar computational cost. The same procedure is used, beginning with trial orbitals expanded in terms of basis functions, then solving the SCF equations. The same methods applied to HF for determining molecular properties can thus be applied to KS theory. Despite these similarities, the advantage of KS theory is that unlike HF, which neglects correlation, it is formally exact.

2.5 Exchange–correlation functionals

Having emphatically derived such a formally exact theory, one might be tempted to assume the problem is solved—we have a working method, which includes correlation and is no more computationally difficult than Hartree–Fock. There is, however, a remaining problem, in that the exact form of the exchange–correlation energy functional $E_{\mathrm{XC}}[\rho]$ (hereafter often simply referred to as "functional" or "E_{XC}" for brevity) is still unknown, and an approximation must still be made. Although the contribution of E_{XC} is relatively small, the quality of the approximation directly affects the accuracy of any calculations. Understandably, finding a universal functional form appropriate for all systems ranging from a single proton to the largest of proteins, in terms of the simple, three-dimensional density, is not a trivial task, and a large part of the last

fifty years of research in DFT has been dedicated to creating and improving approximations to the elusive E_{XC} functional.

Before discussing in detail the common schemes of functional development, we will briefly comment on the extension of DFT to a spin-dependent form, necessary for modelling the effect of a magnetic field on the spins of electrons, and in general improving the approximations in the absence of a magnetic field (in particular for open-shell systems). For a full derivation and discussion, see section 8.1 of Ref. 58.

We consider separately the α-electron and β-electron densities

$$\rho_\alpha(\boldsymbol{r}) = \sum_i^{n_\alpha} |\psi_i(\boldsymbol{r},\alpha)|^2 \,, \tag{2.41}$$

and

$$\rho_\beta(\boldsymbol{r}) = \sum_i^{n_\beta} |\psi_i(\boldsymbol{r},\beta)|^2 \,, \tag{2.42}$$

with the total density given by

$$\rho(\boldsymbol{r}) = \rho_\alpha(\boldsymbol{r}) + \rho_\beta(\boldsymbol{r}) \,; \tag{2.43}$$

n_α and n_β are the number of α- and β-spin electrons, respectively. In the case of closed-shell systems, $\rho_\alpha = \rho_\beta = \rho/2$.

2.5.1 Local density approximation

The local density approximation (LDA), or more properly the local *spin* density approximation (LSDA) in the spin-polarised formalism, is the simplest physically relevant approximation to E_{XC}, and is the form originally demonstrated in Kohn and Sham's seminal paper. Derived from a consideration of the UEG and applying it locally, the functional is split into separate exchange ($E_{\mathrm{X}}[\rho]$) and correlation ($E_{\mathrm{C}}[\rho]$) functionals, where the overall functional is given by $E_{\mathrm{XC}} = E_{\mathrm{X}} + E_{\mathrm{C}}$.

The exchange functional takes the form

$$E_{\mathrm{X}}^{\mathrm{LSDA}}[\rho_\alpha,\rho_\beta] = 2^{1/3} C_{\mathrm{X}} \sum_\sigma \int \rho_\sigma^{4/3}\, d\boldsymbol{r} \,, \quad \text{where} \quad C_{\mathrm{X}} = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} \,, \tag{2.44}$$

and σ is a spin coordinate indicating summation over α and β. This is merely the Dirac approximation of equation (2.7), generalised to a spin-polarised form.

A subtle difference here is that whilst the original TFD arguments used the UEG to model $T[\rho]$ and $V_{\text{ee}}[\rho]$, here we are restricting its use to the smaller, unknown component of KS theory.

For the correlation functional no explicit form is known, however accurate values have been calculated[59] using quantum Monte Carlo. Analytic parameterisations of these values have been generated, the most common being that of Vosko, Wilk and Nusair[60] (VWN).

The LSDA has seen popularity in the field of physics, where its application to large, periodic systems such as bulk metals is relatively successful. For the isolated molecules relevant to chemical problems, however, the localised densities bear very little resemblance to the UEG, and so the approximation quickly breaks down. Several deficiencies are apparent, most notably the tendency to over-bind molecules, and a non-zero correlation energy for one-electron systems.

2.5.2 GENERALISED GRADIENT APPROXIMATIONS

Given the inhomogeneity in the density of a typical molecule, a natural progression to the functional form is to include information about the density gradient. The most common approach is the generalised gradient approximation (GGA), which takes the general form

$$E_{\text{XC}}^{\text{GGA}}[\rho] = \int F(\rho, \nabla\rho(\boldsymbol{r}))\, d\boldsymbol{r}\,, \tag{2.45}$$

which can again be extended to a spin-polarised form in terms of ρ_α and ρ_β. As before, the functional is usually partitioned into separate exchange and correlation components.

The exchange functional is typically expanded in terms of a dimensionless density gradient, a natural example being

$$x(\boldsymbol{r}) = \frac{\nabla\rho(\boldsymbol{r})}{\rho^{4/3}(\boldsymbol{r})}\,, \tag{2.46}$$

such that the form

$$E_{\text{X}}^{\text{GGA}}[\rho] = \int \rho^{4/3}(\boldsymbol{r}) f(x(\boldsymbol{r}))\, d\boldsymbol{r} \tag{2.47}$$

maintains correct coordinate scaling (further discussion of the scaling properties of functionals is given in Chapter 5).

Two distinct approaches to the development of GGA functionals have been widely utilised. The first is to determine a form which, through various parameters, reproduces as many known mathematical properties of the exact exchange and correlation functionals as is practical. Prominent functionals developed in this manner include the Perdew–Wang[61,62] 1991 (PW91) and Perdew–Burke–Ernzerhof[63] (PBE) functionals.

The alternative approach is to determine a parameterised form, where the parameters are empirically fitted to known molecular properties. Notable exchange functionals of this type include Becke's 1986[64] and 1988[65] (B88) functionals, fitted to atomic energies. Most prominently among correlation functionals, the Lee–Yang–Parr[66] (LYP) approximation was derived from Colle–Salvetti[67] calculations on the helium atom.

Both approaches to functional development have seen notable improvements over the LSDA, and each has its advantages. On the one hand, the functionals derived from known mathematical relationships tend to offer more physical insight into the reasons for their success. On the other, empirically derived functionals have tended to show a greater proclivity to achieve good-quality results, although the approximations may tend to break down for systems and properties significantly outside the remit of the empirical fit.

Probably the most famous empirically derived GGA is the combination of Dirac exchange, B88 and LYP, collectively termed BLYP.[68] The immediate success of this functional kick-started the expansion of DFT into the chemical field, whereas it had previously been thought primarily useful to physics. The BLYP functional also provides the foundation for the immensely popular B3LYP, discussed in Section 2.5.4.

Despite showing an almost universal improvement over the LSDA, and giving a reasonable approximation of atomisation energies, ion energetics and local excitation energies, GGA functionals still fail to give quantitative accuracy in a number of areas. Particular failures include the underestimation of reaction barriers, NMR shieldings, and Rydberg and charge–transfer excitation energies.

2.5.3 META-GGAS

Given that the LSDA and the gradient correction can be thought of as the first two terms of a Taylor series, a logical next step would be to introduce

higher-order derivatives of the density. In principle this would involve—as the next term—the introduction of the Laplacian of the density $\nabla^2\rho(\boldsymbol{r})$, however in practice the more numerically stable kinetic energy density $\tau = \sum_i |\nabla\psi_i|^2$ is typically used.

Again, functionals may be derived from theoretical arguments, as with the PKZB[69] and TPSS[70,71] functionals, or from empirical fitting, as with the VSXC functional of Van Voorhis and Scuseria,[72] and the local Minnesota functionals of Truhlar and co-workers, M06-L[73,74] and M11-L.[75] The increased flexibility of the functional form, for little extra computational cost, is a promising concept, however initial implementations did not exhibit any major improvement over conventional GGAs for chemical applications. Recent advances have been made which improve the prospects for these functionals, in particular when combined with the hybrid and range-separated approaches discussed below.[73,76–78]

2.5.4 HYBRID FUNCTIONALS

Consideration of the adiabatic connection provides an alternative approach to improving the functional form: that of incorporating a proportion of exact exchange $E_{\mathrm{X}}^0[\rho]$. Consider a generalisation of the definition of $F[\rho]$ from the constrained search approach, equation (2.24),

$$F_\lambda[\rho] = \min_{\psi\to\rho} \langle\psi|\hat{T} + \lambda\hat{V}_{\mathrm{ee}}|\psi\rangle = \langle\psi_\lambda|\hat{T} + \lambda\hat{V}_{\mathrm{ee}}|\psi_\lambda\rangle \;, \tag{2.48}$$

where λ controls the strength of electron–electron interaction, and ψ_λ is the wavefunction that minimises $\langle\hat{T} + \lambda\hat{V}_{\mathrm{ee}}\rangle$ and gives the exact density. By choosing $\lambda = 0$ for the non-interacting system and $\lambda = 1$ for the fully interacting system we find

$$F_1[\rho] = T[\rho] + V_{\mathrm{ee}}[\rho] \,, \tag{2.49}$$

$$F_0[\rho] = T_{\mathrm{s}}[\rho] \,, \tag{2.50}$$

and, from equation (2.32),

$$E_{\mathrm{XC}}[\rho] = F_1[\rho] - F_0[\rho] - J[\rho] \,. \tag{2.51}$$

If we assume that the functional $F_\lambda[\rho]$ smoothly varies to connect the non-interacting system with the fully interacting system, *via* a series of partially

interacting systems, then

$$E_{\mathrm{XC}}[\rho] = \int_0^1 \frac{\partial F_\lambda[\rho]}{\partial \lambda}\, d\lambda - J[\rho] \tag{2.52}$$

$$= \int_0^1 W_\lambda \, d\lambda \,, \tag{2.53}$$

where

$$W_\lambda = \langle \psi_\lambda | \hat{V}_{\mathrm{ee}} | \psi_\lambda \rangle - J[\rho] \,. \tag{2.54}$$

An appropriate choice of W_λ can therefore be used to approximate the exchange–correlation energy, exploiting the fact that W_0 corresponds exactly to E_{X}^0, whilst W_1 corresponds exactly to $V_{\mathrm{ee}} - J$, which—although the exact form is unknown—can be approximated using existing functionals.

Early considerations by Becke[79] chose W_λ as a linear function of λ, $W_\lambda = a + b\lambda$, where $W_0 = a$ is represented by exact exchange and $W_1 = a + b$ by LSDA to give the *half-and-half* functional BHH (often combined with LYP correlation to give BHLYP),

$$E_{\mathrm{XC}}^{\mathrm{BHH}} = a + \frac{b}{2} = \frac{1}{2} E_{\mathrm{X}}^0 + \frac{1}{2} U_{\mathrm{XC}}^{\mathrm{LSDA}} \,. \tag{2.55}$$

where $U_{\mathrm{XC}}^{\mathrm{LSDA}}$ is the potential energy contribution to the LSDA exchange–correlation energy. Even with such a simple approximation, Becke saw promising improvements to thermochemical properties, and established the powerful observation that exchange–correlation functionals could be improved by incorporating a fraction of exact exchange E_{X}^0 into the functional form.

By introducing a number of semi-empirical parameters, Becke subsequently proposed the B3PW91 functional,[80]

$$E_{\mathrm{XC}}^{\mathrm{B3PW91}} = (1 - A) E_{\mathrm{X}}^{\mathrm{Dirac}} + A E_{\mathrm{X}}^0 + B \Delta E_{\mathrm{X}}^{\mathrm{B88}} + E_{\mathrm{C}}^{\mathrm{VWN}} + C \Delta E_{\mathrm{C}}^{\mathrm{PW91}} \,, \tag{2.56}$$

with optimal parameters $A = 0.20$, $B = 0.72$ and $C = 0.81$. A slightly modified version proposed by Stephens *et al.*[81] replaced the PW91 correlation with LYP, to give B3LYP,

$$E_{\mathrm{XC}}^{\mathrm{B3LYP}} = (1 - A) E_{\mathrm{X}}^{\mathrm{Dirac}} + A E_{\mathrm{X}}^0 + B \Delta E_{\mathrm{X}}^{\mathrm{B88}} + (1 - C) E_{\mathrm{C}}^{\mathrm{VWN}} + C E_{\mathrm{C}}^{\mathrm{LYP}} \,, \tag{2.57}$$

which has seen unparalleled success in the field of chemistry and remains a tremendously popular method.

The general form of a hybrid functional is often denoted

$$E_{\mathrm{XC}}^{\mathrm{hybrid}}[\rho] = \int F(\rho, \nabla\rho)\, d\boldsymbol{r} + \xi E_{\mathrm{X}}^{0}\,, \tag{2.58}$$

where the fraction of exact exchange is controlled by the parameter ξ. In general, functionals of this form are implemented within the generalised Kohn–Sham (GKS) framework,[82] where the energy is minimised with respect to the orbitals (as opposed to the density, as in strict KS theory).

There is an important distinction to be made between this method of incorporating exact exchange into the overall DFT exchange–correlation functional, and that of simply including DFT correlation as a correction to exact exchange. The latter option in general fails to include the non-dynamic correlation absent from exact exchange, but present in local density exchange functionals. Becke and Johnson have investigated additions to exact exchange which model both dynamic and non-dynamic correlation, with the B05 and DF07 functionals.[83–87]

The widespread popularity of B3LYP stems largely from its improvement over GGAs for computing thermochemical properties, making it widely applicable to many aspects of chemistry. However, although the failings of GGAs in other areas—such as reaction barriers, Rydberg and charge–transfer excitation energies—are reduced somewhat, the errors remain too high for quantitative use, and NMR shieldings tend to be less accurate than for GGAs.

A vast array of alternative hybrid functionals exist, derived from both theoretical and semi-empirical arguments. The PBE0 functional[88,89] combines the parameter-free PBE with and zero-parameter arguments for the partitioning of E_{X}^{0} and $E_{\mathrm{X}}^{\mathrm{GGA}}$ following Becke[90] and Perdew *et al.*,[91] to create a hybrid functional free from any semi-empirical parameters. Conversely a series of functionals based on the Becke 1997 expansion[92] have been proposed, comprising B97-1,[93] B97-2,[94] and B97-3.[95]

Finally, we note the MCY functionals of Mori-Sánchez, Cohen, and Yang,[96–98] who considered an alternative [1,1]-Padé form of the adiabatic connection previously proposed by Ernzerhof.[99] A key feature of this series of functionals is that by construction the functionals are free from one-electron (although not necessarily many-electron) self-interaction. The importance of the self-interaction problem is discussed in Section 2.7.1 and Chapter 3.

2.5.5 Range-separated functionals

Becke's original half-and-half approximation showed that a larger proportion of E_{X}^{0} than is typically seen in general hybrid functionals is actually beneficial in some cases. Specifically, long-range properties can be significantly improved, but at the detriment to the quality of the short-range properties exhibited by B3LYP *et al.* This is not altogether surprising if one considers that correlation effects will be strongest at shorter range, so the exact exchange term (which neglects correlation) becomes increasingly accurate at long range.

In recent years, much work has been directed towards *range-separated* functionals,[100,100–118] where the functional form is partitioned into short- and long-range components, with each modelled differently. Most commonly this takes the form of incorporating a greater proportion of exact exchange at long range than at short, the idea being to maintain the quality of the short-range properties shown by conventional hybrids, whilst improving the long-range properties.

The precise balance between the long- and short-range components can have a profound effect on the behaviour of the functional, and this variation, along with its benefits, is the subject of Chapters 3 and 4. A full discussion of the partitioning schemes is contained therein.

2.6 Time-dependent DFT

The Hohenberg–Kohn theorems rely on the use of the variational principle, and so only the ground state of a system is accessible. In order to probe excited states and related properties the theory must be extended, in the form of time-dependent DFT (TDDFT).

Analogous to the Hohenberg–Kohn proofs for ground-state DFT, TDDFT is founded upon the Runge–Gross[119] theorem, which states that the time-dependent external potential $v(\boldsymbol{r},t)$, and thus the time-dependent wavefunction $\psi(\boldsymbol{r},t)$, is uniquely determined (the former to within a spatially constant function and the latter to within a time-dependent phase-factor) by the exact time-dependent density $\rho(\boldsymbol{r},t)$. Similarly, a variational principle can be established involving the action integral.

The Kohn–Sham equations can then be generalised to the time-dependent formalism (TDKS),[120] as

$$\hat{F}^{\text{KS}}\varphi_i(\boldsymbol{r},t) = i\frac{\partial}{\partial t}\varphi_i(\boldsymbol{r},t)\,, \tag{2.59}$$

where $\hat{F}^{\text{KS}}$ is the time-dependent KS operator. Equation (2.59) can be represented in matrix-element notation as

$$\sum_q \left(F_{pq}P_{qr} - P_{pq}F_{qr}\right) = i\frac{\partial}{\partial t}P_{pr}\,, \tag{2.60}$$

where the elements of the density matrix P_{pr} are related to the (time-dependent) density by

$$\rho(\boldsymbol{r},t) = \sum_{p,q} c_p(t)c_q(t)\varphi_p(\boldsymbol{r})\varphi_q(\boldsymbol{r}) \tag{2.61}$$

$$= \sum_{p,q} P_{pq}(t)\varphi_p(\boldsymbol{r})\varphi_q(\boldsymbol{r})\,, \tag{2.62}$$

and $p, q, \ldots$ are generic indices denoting any orbital ($i, j, \ldots$ refer to occupied and $a, b, \ldots$ to virtual orbitals).

In principle, the TDKS expression involves a time-dependent exchange–correlation functional; however through the adiabatic approximation, which assumes the density varies slowly with time, it can be approximated in terms of the time-*independent* XC functional (evaluated with the time-dependent density).

2.6.1 LINEAR-RESPONSE TDDFT

Most commonly, a linear-response formalism of TDDFT is employed,[120–123] where excitation energies are determined by considering the response of the ground state density to the small perturbation by a time-dependent electric field, such that

$$P_{pq} = P_{pq}^{(0)} + P_{pq}^{(1)}\,, \tag{2.63}$$

$$F_{pq} = F_{pq}^{(0)} + F_{pq}^{(1)}\,, \tag{2.64}$$

which by substitution into equation (2.60) (and collection of the first-order terms) yield

$$\sum_q \left(F_{pq}^{(0)}P_{qr}^{(1)} - P_{pq}^{(1)}F_{qr}^{(0)} + F_{pq}^{(1)}P_{qr}^{(0)} - P_{pq}^{(0)}F_{qr}^{(1)}\right) = i\frac{\partial}{\partial t}P_{pr}^{(1)}\,. \tag{2.65}$$

The perturbation to the KS Hamiltonian matrix $\boldsymbol{F}$ consists of two terms: the contribution from the applied electric field (of frequency ω)

$$g_{pq} = \frac{1}{2}\left(f_{pq}\mathrm{e}^{-i\omega t} + f^{*}_{qp}\mathrm{e}^{i\omega t}\right) , \tag{2.66}$$

where f_{pq} is a one-electron operator describing the perturbation, and the response of the two-electron part of $\boldsymbol{F}$ to changes in the density matrix $\boldsymbol{P}$,

$$\Delta F^{(0)}_{pq} = \sum_{st} \frac{\partial F^{(0)}_{pq}}{\partial P_{st}} P^{(1)}_{st} . \tag{2.67}$$

The overall first-order change in $\boldsymbol{F}$ is the sum of these two terms,

$$F^{(1)}_{pq} = g_{pq} + \Delta F^{(0)}_{pq} . \tag{2.68}$$

The first-order perturbation to the density matrix $\boldsymbol{P}$ itself is given by

$$P^{(1)}_{pq} = \frac{1}{2}\left(d_{pq}\mathrm{e}^{-i\omega t} + d^{*}_{qp}\mathrm{e}^{i\omega t}\right) , \tag{2.69}$$

where d_{pq} represent elements of the perturbation density matrix.

If we substitute equations (2.68) and (2.69) into equation (2.65) and compare the terms multiplying $\mathrm{e}^{-\omega t}$, we obtain the expression

$$\begin{aligned}\sum_{q}\Bigg[F^{(0)}_{pq}d_{qr} - d_{qr}F^{(0)}_{qr} + \left(f_{pq} + \sum_{st}\frac{\partial F^{(0)}_{pq}}{\partial P_{st}}d_{st}\right)P^{(0)}_{qr} \\ - P^{(0)}_{pq}\left(f_{qr} + \sum_{st}\frac{\partial F^{(0)}_{qr}}{\partial P_{st}}d_{st}\right)\Bigg] = \omega d_{pr} ;\end{aligned} \tag{2.70}$$

similarly the complex conjugate of equation (2.70) is obtained by comparing the terms multiplying $\mathrm{e}^{i\omega t}$.

From the idempotency condition

$$\sum_{q} P^{(0)}_{pq}P^{(1)}_{qr} = P^{(0)}_{pr} , \tag{2.71}$$

we can identify the first-order change in $\boldsymbol{P}$ as

$$\sum_{q}\left(P^{(0)}_{pq}P^{(1)}_{qr} + P^{(1)}_{pq}P^{(0)}_{qr}\right) = P^{(1)}_{pr} , \tag{2.72}$$

and so only the occupied–virtual and virtual–occupied elements of the matrix d_{pq} (d_{ia} and d_{ai}, respectively) are non-zero. Given that the unperturbed KS Hamiltonian and density matrices are diagonal, we obtain

$$F_{aa}^{(0)} d_{ai} - d_{ai} F_{ii}^{(0)} + \left[f_{ai} + \sum_{b,j} \left(\frac{\partial F_{ai}}{\partial P_{bj}} d_{bj} + \frac{\partial F_{ai}}{\partial P_{jb}} d_{jb} \right) \right] P_{ii}^{(0)} = \omega d_{ai} \,, \tag{2.73}$$

and

$$F_{ii}^{(0)} d_{ia} - d_{ia} F_{aa}^{(0)} - P_{ii}^{(0)} \left[f_{ia} + \sum_{b,j} \left(\frac{\partial F_{ia}}{\partial P_{bj}} d_{bj} + \frac{\partial F_{ia}}{\partial P_{jb}} d_{jb} \right) \right] = \omega d_{ia} \,, \tag{2.74}$$

In order to follow conventional notation, we define $d_{ai} = x_{ai}$ and $d_{ia} = y_{ai}$. In the zero-frequency limit, where an infinitesimal perturbation is assumed ($f_{ia} = f_{ai} = 0$), we can rewrite equations (2.73) and (2.74) as the non-Hermitian eigenvalue equation

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B^*} & \boldsymbol{A^*} \end{pmatrix} \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix} = \omega \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix} \,, \tag{2.75}$$

where we have made use of the relations $F_{pp}^{(0)} = \varepsilon_p$ and $P_{ii}^{(0)} = 1$. Equation (2.75) is then solved to determine the excitation energies ω.

The elements of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are given (for local or GGA functionals) by

$$A_{ia,jb} = \delta_{ij}\delta_{ab}(\varepsilon_a - \varepsilon_i) + (ia|jb) + (ia|f_{\mathrm{XC}}|jb) \,, \tag{2.76}$$

and

$$B_{ia,jb} = (ia|bj) + (ia|f_{\mathrm{XC}}|bj) \,, \tag{2.77}$$

where we switch to the Mulliken "charge-cloud" notation for two-electron integrals. The quantity f_{XC} is termed the exchange–correlation kernel, and is the second functional derivative of E_{XC},

$$(ia|f_{\mathrm{XC}}|jb) = \iint d\boldsymbol{r}\, d\boldsymbol{r'}\, \varphi_i^*(\boldsymbol{r}) \varphi_a(\boldsymbol{r}) \frac{\delta^2 E_{\mathrm{XC}}}{\delta\rho(\boldsymbol{r})\delta\rho(\boldsymbol{r'})} \varphi_b^*(\boldsymbol{r'}) \varphi_j(\boldsymbol{r'}) \,. \tag{2.78}$$

Note that additional terms are present in equations (2.76) and (2.77) for hybrid and range-separated functionals.

2.7 Specific failures of conventional DFT

Despite the plethora of functionals now available, and the many years of research dedicated to improve them, there remain several challenges in DFT functional development, owing to a number of failures of virtually all conventional approximate functionals. A comprehensive review by Cohen *et al.*[124] gives a detailed account of the major hurdles remaining in DFT; this thesis will focus in large part on the delocalisation error, and to some extent the asymptotic behaviour of the exchange–correlation potential.

2.7.1 Delocalisation error

A number of failings that persist in approximate DFT—including the underestimation of reaction barriers,[125–130] band gaps,[131,132] energies of dissociating ions,[133–136] and charge–transfer excitation energies,[137]–are symptomatic of a common cause: delocalisation error.[133,134,138] The delocalisation error can in part be traced to the unphysical interaction of an electron with itself, which conventional functionals fail to cancel, however the consequences extend far beyond a simple one-electron problem.

The most intuitive illustration of the problem can be seen by stretching $H_2{}^+$, which—as a one-electron system—can be described exactly by Hartree–Fock theory. At equilibrium, this simple system of a single electron shared between two centres is well modelled by conventional exchange–correlation functionals. However, as the bond is stretched and the system approaches two separated H nuclei, each formally with half an electron, the energy is significantly underestimated by approximate DFT.

The traditional understanding of this incorrect behaviour is the failure of approximate functionals to cancel the electron–electron interaction terms $J[\rho]$ and $E_{\text{XC}}[\rho]$, which should sum to zero for a single electron. The resulting *self-interaction error* (SIE) has been widely discussed in the literature,[97,133,139–143] and is defined as the sum of the above two terms,

$$\text{SIE} = J[\rho] + E_{\text{XC}}[\rho]\,. \tag{2.79}$$

A number of efforts have been made to correct this SIE. Perdew and Zunger[144] presented a correction term that eliminates the one-electron self-interaction

terms, though this brought other complications, such as detrimental effects to atomisation energies and equilibrium properties,[145–147] and a lack of invariance with respect to unitary transformation of the occupied orbitals.[148] Becke's B05,[83] and the MCY functionals of Mori-Sánchez, Cohen, and Yang,[96,97] are also free from the one-electron SIE whilst improving thermochemistry and reaction barriers.

However, the problem extends beyond a simple one-electron cancellation. Zhang and Yang[143] demonstrated that the SIE will increase for fractionally charged systems (as demonstrated by the stretching of H_2^+), and that even if the SIE is eliminated for a one-electron system it will—without proper consideration of the scaling properties of the functional—still exist for systems with $0 \leqslant N < 1$ electrons. Though the SIE is easiest to define and analyse as a one-electron problem, its consequences extend to many-electron systems. The extension to the *many-electron* self-interaction error (MESIE)[97,133,134,139–142] is much more difficult to conceptualise, however the key observation is that functionals incorporating corrections to the one-electron SIE still exhibit erroneous behaviour associated with its presence, namely the incorrect lowering of energy (*i.e.* stabilisation) of fractional charges.

The upshot of the error is that approximate exchange–correlation functionals tend to incorrectly over-stabilise (and hence favour) systems that locally exhibit fractional charges. Put another way, these functionals tend to over-delocalise the charge distribution in order to (incorrectly) lower the energy, and so the term *delocalisation error* is used to capture the physical manifestation of the underlying problem.

E vs N plots

Plotting the energy E of a system with respect to a fractional variation in the number of electrons N (abbreviated to E *vs* N) can model the extent of the error, and has generated significant interest in recent years.[97,131–134,138,139,142,149–160] DFT calculations on a system with an arbitrary fractional number of electrons can be carried out by explicitly choosing a fractional occupation number for the HOMO in the one-particle density matrix.[133,161]

The exact behaviour was determined by Perdew *et al.*,[163] who used a zero-temperature ensemble approach to model an open system free to exchange

electrons. Consider a system with $N = N_0 + \delta$ electrons, for integer N_0 and $0 \leqslant \delta \leqslant 1$. The quantum mechanical ensemble is described by a linear combination of pure states—in this case ψ_{N_0} and ψ_{N_0+1}—weighted by their respective probabilities $(1 - \delta)$ and δ. The authors demonstrated that the energy is given by

$$(1 - \delta)E_{N_0} + \delta E_{N_0+1}\,, \tag{2.80}$$

and as such the exact E *vs* N takes the form of a series of piecewise linear segments, with discontinuities in the gradient at integer N. An alternative proof, without invoking the ensemble approach, was given by Yang *et al.*[164] in the limit of dissociation.

This discontinuity in the derivative of the energy manifests as a discontinuity in the potential as N passes through an integer: the potential on the electron-deficient side vanishes asymptotically, whereas the potential on the electron-abundant side is identical in shape, yet shifted by a constant known as the derivative discontinuity Δ_{XC}.[165] By definition, for a system with integer $N = N_0$ electrons, the gradient of the slope on the electron-deficient side of N_0 is equal to the vertical ionisation potential I_0, whereas the gradient on the electron-abundant side is the vertical electron affinity A_0.

In Hartree–Fock theory, one can relate the eigenvalues of occupied orbitals to corresponding ionisation potentials through Koopmans' theorem,[166] which states that the negative of the occupied orbital energy is equal to the ionisation potential due to removal of the same electron, when the orbitals are frozen (the condition becomes an approximation when the orbitals are allowed to relax). Within DFT, Janak's theorem[167] provides an analogous exact condition for the frontier orbitals, stating that the change in energy with respect to the occupation number of an orbital (*i.e.* $\partial E/\partial n = \partial E/\partial N$) is equal to the eigenvalue of that orbital. Combining this with the exact linearity condition equation (2.80), we see that

$$\varepsilon_{N_0}(N_0 - f) = -I_0\,, \tag{2.81}$$

and

$$\varepsilon_{N_0+1}(N_0 + f) = -A_0\,, \tag{2.82}$$

where $0 < f < 1$.

In the limit $f \to 0$, these orbital energies correspond to the highest-occupied (HOMO) and lowest-unoccupied (LUMO) molecular orbital energies for the N_0-electron system, evaluated with the exchange–correlation potentials on the electron-deficient and electron-abundant sides, respectively, of N_0. Denoting these orbital energies $\varepsilon_{\mathrm{H}}^{-}$ and $\varepsilon_{\mathrm{L}}^{+}$, we have

$$\varepsilon_{\mathrm{H}}^{-} = -I_0\,, \tag{2.83}$$

and

$$\varepsilon_{\mathrm{L}}^{+} = -A_0\,, \tag{2.84}$$

giving what we will call (exact) Koopmans conditions in DFT.

In the limit $f \to 1$, equation (2.81) corresponds to $\varepsilon_{\mathrm{L}}^{+}$ of the $(N_0 - 1)$-electron system, and by definition $A_0^{(N_0-1)} \equiv I_0^{(N_0)}$. Similarly, equation (2.82) corresponds to $\varepsilon_{\mathrm{H}}^{-}$ of the $(N_0 + 1)$-electron system, where $I_0^{(N_0+1)} \equiv A_0^{(N_0)}$. It follows that $\varepsilon_{\mathrm{L}}^{+}(N_0 - 1)$ must equal $\varepsilon_{\mathrm{H}}^{-}(N_0)$ and $\varepsilon_{\mathrm{L}}^{+}(N_0)$ must equal $\varepsilon_{\mathrm{H}}^{-}(N_0 + 1)$, with the orbital energy remaining constant between each pair of integer N. The use of the $\pm$ superscript to denote the side of the integer is vital, because the exact exchange–correlation potential jumps discontinuously as the integer is crossed, meaning a given orbital energy also jumps by the same amount.

Two inter-related deficiencies are characteristic of approximate explicit density functionals. Firstly, the delocalisation error produces unphysical curvature in E *vs* N, due to the lowering of the energy at fractional N. Secondly, there is no discontinuity in the potential, and so at best they can average over it. A recent paper by Stein *et al.*[154] discusses the intrinsic link between the two problems.

These deficiencies have serious repercussions. In practical calculations, using approximate exchange–correlation functionals within the usual generalised Kohn–Sham approach,[82] $\partial E/\partial N$ is again equal to the orbital energy[131] and so the values of $\partial E/\partial N$ on the $f \to 0$ electron-deficient and electron-abundant sides of N_0 equal the HOMO energy ε_{H} and LUMO energy ε_{L} of the N_0-electron system, respectively. Explicit density functionals do not satisfy the conditions in equations (2.81) and (2.82) due to the inherent curvature in E *vs* N associated with the delocalisation error, leading to HOMO energies much greater than $-I_0$ and LUMO energies much lower than $-A_0$.

Conversely, Hartree–Fock—aside from a one-electron system for which it is exact—incorrectly *raises* the energy for fractional charges, creating a *localisation error* and concave E *vs* N curvature. This causes a useful partial cancellation of errors for hybrid functionals where (depending on the exact fraction of exact exchange included) problematic quantities can be improved, although it is far from a rigorous, system-independent solution.

Manifestation of the error

Ruzsinszky *et al.*[140] have shown that neutral molecules may incorrectly dissociate to fragments containing fractional charges, when modelled with functionals that suffer from a large delocalisation error. NaCl, for example, dissociates to approximately $Na^{0.4+}$ and $Cl^{0.4-}$ when computed with PBE. This is an artifact of the unphysical lowering of energy of the fractionally charged systems compared to the integer case. Incorporating a self-interaction correction scheme lessens this spurious behaviour by reducing—although not eliminating—the delocalisation error.

Peach *et al.*[168] and Heaton-Burgess and Yang[169] have noted that B3LYP can give a poor description of bond length alternation (BLA)—a structural manifestation of the degree of electron-delocalisation in conjugated π-systems. Explicit density functionals, and hybrids with a lower proportion of exact exchange such as B3LYP, tend to bias the system towards greater delocalisation and hence underestimate the BLA. Both studies show that a larger, more accurate BLA is predicted by the CAM-B3LYP range-separated functional, and Peach *et al.* draw the analogy to similarly improved results given by BHLYP—a fixed hybrid with a larger proportion of exact exchange (although CAM-B3LYP remains preferable due to better applicability as an all-round functional).

Diels–Alder reactions are another good representative example of a problem caused by the delocalisation error.[130,170] These pericyclic reactions, which proceed through a highly delocalised transition state, are very sensitive to the choice of functional. A GGA such as BLYP, with a large delocalisation error, dramatically over-stabilises this delocalised transition state and so underestimates the reaction barrier, whereas including and increasing a proportion of exact exchange with B3LYP and BHLYP systematically reduces this error.

Poater *et al.*[171] have examined in detail the behaviour of the ground and low-

lying electronic states of $Cu^{2+}-H_2O$, as the proportion of exact exchange E_X^0 in hybrid functionals is varied. Higher proportions of E_X^0 predict a ground state of $^2A_1$ in a C_{2v} geometry, in agreement with CCSD(T) predictions. As E_X^0 is reduced, the $^2B_1$ state becomes more stable in C_{2v}, whilst the ground state becomes C_s ($^2A'$). This is rationalised by examination of charge and spin delocalisation—the states which are over-stabilised by, in particular, BLYP and B3LYP, with lower proportions of E_X^0, provide a more delocalised distribution of the electron density, causing erroneous stabilisation by these functionals. Rios-Font *et al.*[172] extend this work to a wider range of systems, studying $Cu^{2+}-(H_2O)_n$ complexes for $n = 1$–6. Again, BLYP and B3LYP tend to predict lower-symmetry structures with more-delocalised electron densities. The admixture of more E_X^0 gives an improved description.

The examples above are far from exhaustive, and it is clear that the delocalisation error is a severe failing in approximate DFT, with far-reaching consequences. It is unsurprising then, that a large amount of recent research has focused on its reduction and elimination.

Reducing delocalisation error by approximately enforcing linearity

A number of approaches can be used to reduce the delocalisation error by imposing near-linear E *vs* N behaviour. Vydrov *et al.*[139] showed that the MESIE was significantly reduced by applying the PZ self-interaction correction;[144] see also Refs 148, 173, and 174. The MCY3 and rCAM-B3LYP functionals[97] were specifically designed to achieve near-linear behaviour, and have shown some success.[169,170] Zheng *et al.*[175] proposed a non-empirical scaling correction to largely restore linearity, which was later extended[176] to properly account for orbital relaxation effects. Although the scaling correction applies to systems with explicitly fractional N, it does not affect integer-N systems with locally fractional regions (such as the case of stretched H_2^+). A recent extension to the scheme[177] introduces a *local* scaling correction to counter this deficiency. Recently, Kraisler and Kronik[178] demonstrated that the exact Koopmans ionisation condition could be largely restored using an ensemble treatment.

Other groups,[172,179,180] in keeping with above observations, have noted that increasing the proportion of exact exchange in hybrid functionals can improve erroneous results associated with the delocalisation error, with functionals such

as BHLYP performing better despite their poor treatment of thermochemical properties. This improved behaviour can be visualised as the result of error-cancellation between the delocalisation of DFT and the localisation of HF. "Local hybrid" functionals,[181–184] can improve the flexibility of the admixture, providing the desired compromise between the components without as much detriment to the thermochemistry. The balance of components in range-separated hybrids will be addressed in Chapter 3.

Static correlation error

Although beyond the scope of this thesis, it would be remiss not to mention the related problem of the static correlation error (SCE).[185,186] Whilst the delocalisation error is characterised by the incorrect treatment of fractional charges, the SCE instead arises due to fractional spins. Such systems are, again, subject to an energy constancy condition where—for the exact functional—a system with δ α-spin electrons and $(1-\delta)$ β-spin electrons has an identical energy for any δ $(0 \leqslant \delta \leqslant 1)$. The failure of approximate density functionals to satisfy this condition is the cause of well-known failures in strongly correlated systems, illustrated by the overestimation of the energy in the simple case of dissociating H_2.[185,187–189]

2.7.2 Exchange–correlation potential

A second important failing of typical approximate exchange–correlation functionals is that their functional derivative—the exchange–correlation potential—exhibits incorrect long-range behaviour. We have already established that approximate functionals fail to exhibit the discontinuity in the potential when passing through an integer N, however an additional property of the exact $v_{\text{XC}}(\boldsymbol{r})$ is that it should vanish asymptotically as $-1/r$ on the electron-deficient side, and tend to $-1/r + \Delta_{\text{XC}}$ on the electron-abundant side.

As discussed in Section 2.7.1, GGA functionals approximately average over the discontinuity, but only in regions of appreciable electron density. In asymptotic regions the averaged potential should tend to $-1/r+\Delta_{\text{XC}}/2$, however for a typical GGA $v_{\text{XC}}(\boldsymbol{r})$ incorrectly vanishes, and at an exponential rate—*i.e.* faster than $-1/r$. Since the core region—which describes the occupied

orbitals—mimics the averaged potential, these orbitals are shifted up relative to the virtual orbitals which are described by the quickly vanishing asymptotic region.

This incorrect description causes Rydberg excitation energies to be greatly underestimated,[190] and so exhibiting the correct asymptotic behaviour is another key desirable condition for any functional. The asymptotic correction[191] scheme has been developed to constrain $v_{\mathrm{XC}}(\boldsymbol{r})$ to obey the limit

$$\lim_{r\to\infty} v_{\mathrm{XC}}(\boldsymbol{r}) = -\frac{1}{r} + I + \varepsilon_{\mathrm{H}}\,, \tag{2.85}$$

where the derivative discontinuity has been approximated as twice the sum of the ionisation potential and the energy of the HOMO.[149,162,192,193] Variations to the asymptotic correction have been proposed,[194,195] along with extensions to hybrid[196] and range-separated[197] functionals. Potentials derived from methods such as this, however, no longer correspond to functional derivatives of the energy, and so are limited in their application. More desirable is a functional form whose functional derivative inherently satisfies the asymptotic condition.

2.7.3 NOVEL APPROACHES TO CORRECTING DEFICIENCIES IN APPROXIMATE DENSITY FUNCTIONALS

Having framed the problems of delocalisation error and the incorrect asymptotic XC potential in the context of the underlying theory, the remainder of this thesis focuses on the analysis and development of novel methods that go some way towards counteracting them.

In Chapter 3 we examine the system-dependent tuning of the proportion of exact exchange in range-separated hybrid functionals, in order to approximately satisfy Koopmans and related conditions. The procedure is a particularly prominent extension of the cancellation between DFT delocalisation and HF localisation errors that is growing in popularity, and has been shown to offer a tangible approach to computing orbital energy differences, which would otherwise be greatly hindered by the delocalisation error. We present a systematic analysis of this tuning method, by examining a range of tuning criteria. By relating the apparent success of the tuned functionals to the underlying quantities involved in the tuning, and to E *vs* N plots for a representative system,

we provide an insight into the nature of the error cancellation and the inherent challenges in attempting to enforce linearity on the functional forms.

In Chapter 4 we extend the analysis by considering the effect of functional tuning on the derivative of the density with respect to N, in the form of the Fukui function. In particular, we examine whether functionals tuned to give near-linearity in E *vs* N are similarly optimal for the density.

Ultimately, the goal is to find a functional form inherently free from the delocalisation error. In Chapter 5, we present a completely new approach, based on scaling relations, for developing a functional that approximately recovers the one of the exact Koopmans conditions associated with exact E *vs* N linearity. The approach has the added advantage that it provides a simple mechanism for recovering the exact asymptotic XC potential. Despite a deceptively simple mathematical form, the success of our initial results show a great deal of promise, and our hope is to inspire a different approach to functional development, leading to the elimination of the delocalisation error and other prominent deficiencies in approximate DFT.

3

Tuned range-separated hybrid functionals

In this chapter, we address a recently proposed scheme to tune range-separated hybrid (RSH) functionals in order to reduce the effects of the delocalisation error. Firstly, the theory of RSH functionals is discussed, and a range of typical functionals compared. A method for tuning these functionals in order to reproduce linearity conditions is then introduced, and a systematic assessment of a range of tuning norms is presented. Finally, the impact of these tuning methods on the *E vs N* behaviour of the functional is comprehensively discussed for a model system, highlighting the role of error-cancellation in the procedure.

As outlined in Chapter 2, range-separated hybrid (RSH) functionals expand upon the concept of hybrid functionals by partitioning the functional form into long- and short-range terms, and handling them differently. Although it adds an additional layer of complexity to the functional form, this powerful concept has numerous advantages and has led to great improvements in typically error-prone calculations.

Consider a global hybrid functional such as B3LYP. Admixture of a certain proportion of exact exchange, E_{X}^{0}, into the DFT functional has a beneficial effect in mitigating the errors in the exchange–correlation approximation, without negating the benefit of the non-dynamic correlation inherent to density-

based exchange functionals. The precise proportion is a compromise, often determined by semi-empirically averaging over a range of systems and properties to obtain the most effective balance of the two components. This, however, is an over-simplified picture, and the "optimal" balance is not universally applicable. B3LYP, for example, gives a reasonable to good estimation of atomisation energies, ionisation potentials, and bond lengths, but breaks down for reaction barriers, polarisabilities and excitation energies (in particular Rydberg excitations).[95,198]

These failures typically arise either directly from an insufficient proportion of E_{X}^{0}, or from the incorrect asymptotic behaviour of the exchange–correlation potential (which would not be exhibited by a local potential associated with pure E_{X}^{0}). Clearly then, the fixed proportion of exact exchange in a conventional hybrid is insufficient for a universally applicable functional.

We now consider a more complex partitioning between the approximate-DFT and exact exchange components, which involves modifying the treatment of exchange to depend on the inter-electron distance r_{12}. For finite molecular systems, this approach, commonly known as range-separation, long-range correction, or Coulomb attenuation, usually involves including a small proportion of exact exchange at short range (small r_{12}), increasing smoothly to a larger proportion at long range (large r_{12}). The remainder of the exchange term is treated with conventional DFT approximations. For the purposes of this work, we will describe this class of functionals as *range-separated hybrids* (RSH), and use *long-range correction* and *Coulomb attenuation* to describe two specific partitioning schemes, defined in Sections 3.1.1 and 3.1.2, respectively.

In general, the range-separated approach can provide a good compromise between the DFT and exact-exchange components, improving the overall applicability of the functional. The effect of the precise balance of the components, through parameterisation of the partitioning function, is the focus of Chapters 3 and 4. Firstly, we introduce the common formulations for partitioning schemes, which will determine the underlying functional forms used throughout these chapters.

3.1 Functional forms

3.1.1 Long-range correction

Early long-range correction schemes proposed[100,101] a partitioning of the r_{12}^{-1} operator using the error function, to give

$$\frac{1}{r_{12}} \equiv \frac{\mathrm{erf}(\mu r_{12})}{r_{12}} + \frac{1-\mathrm{erf}(\mu r_{12})}{r_{12}}\,. \tag{3.1}$$

The first term, governing the long-range behaviour, was used to evaluate a modified orbital expression for exact exchange, whereas the second term (short-range part) was used to evaluate DFT approximations—initially the LDA[100,102,103] and extended to GGA by Hirao and co-workers.[104–110,199] In this way the relative proportion of exact-exchange and DFT components is dependent upon the inter-electron distance r_{12}. The parameter μ, denoted the range-separation parameter, determines the ratio of the two parts at a given r_{12}, *i.e.* the rate at which the partitioning varies with inter-electron distance. As a result, the proportion of exact exchange varies smoothly from 0 % to 100 % as r_{12} increases, at a rate dependent on μ. An optimal value of $\mu = 0.33\,{a_0}^{-1}$ was determined by Tawada *et al.*[106] for the long-range–corrected GGA.

This technique, initially termed a *long-range correction scheme* (denoted LC in this work) was shown to produce significant improvements for calculation of long-range properties: errors were reduced in van der Waals interactions,[105] polarisabilities of π-conjugated chains,[104] and both Rydberg and charge–transfer excitation energies.[106] However, the functional failed to reproduce the accuracy of conventional functionals for short-range properties such as atomisation energies, due to little or no exact exchange included at short inter-electron distances. In other words, at short range the functional resembles a pure GGA and much of the benefit of hybrid functionals is lost.

3.1.2 Coulomb attenuation

Expanding upon the idea of long–range correction, Yanai *et al.*[111] generalised the partitioning formula, introducing additional parameters to define lower and upper limits for the proportion of E_{x}^{0}. This *Coulomb-attenuated* (CAM) form is

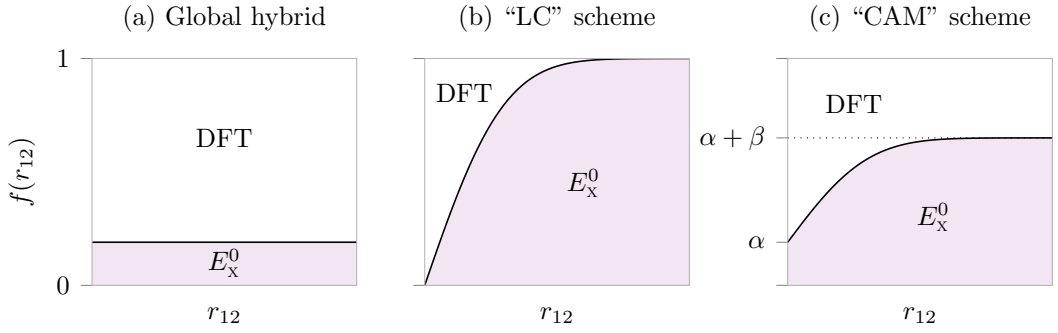

Figure 3.1: Schematic illustration of the partitioning of DFT and exact-exchange in hybrid and range-separated functionals.

given by

$$\frac{1}{r_{12}} \equiv \frac{\alpha + \beta \operatorname{erf}(\mu r_{12})}{r_{12}} + \frac{1 - (\alpha + \beta \operatorname{erf}(\mu r_{12}))}{r_{12}}, \tag{3.2}$$

where α controls the proportion of E_{X}^{0} at short range ($r_{12} = 0$), increasing to $\alpha + \beta$ at long range ($r_{12} \to \infty$); again, the range-separation parameter μ determines the rate at which exact exchange is incorporated.

Applying this CAM scheme to Becke's famous B3LYP hybrid, Yanai *et al.*[111] proposed parameters again fitted to known data, producing the CAM-B3LYP functional. The optimal parameters were determined to be $\alpha = 0.19$, $\alpha + \beta = 0.65$ and $\mu = 0.33\, {a_0}^{-1}$. This gives a short-range proportion of exact exchange very similar to the 0.2 of B3LYP, whilst improving the long-range behaviour of the functional,[198] however the optimal parameters are not universal.[200]

The schematic diagram in Figure 3.1 illustrates the partitioning of the DFT and exact-exchange components for the different classes of hybrid functional. In the range-separated cases, choosing a greater value of μ would increase the steepness of the initial slope.

3.1.3 Alternative partitioning schemes

Other partitioning schemes have been considered, for example by Toulouse *et al.*,[112] and Baer and Neuhauser.[201] We also mention the screened approach of Scuseria and co-workers,[113–117] in which the opposite partitioning is used, with exact exchange dominating the short-range components and excluded at long range. Such a scheme can have computational advantages for extended systems

such as solids, where polarisation effects screen the long-range interactions between electrons and the asymptotic behaviour of the exchange–correlation potential is less important. Neglecting the computationally demanding calculation of exact exchange at long range can therefore increase the tractability without diminishing the accuracy of the functional. In discrete molecular systems, however, the accurate asymptotic description of the potential becomes vital, and so increased long-range exact exchange is more appropriate. For a full discussion, see Ref. 118.

3.2 Tuning functionals

The idea of tuning the parameters of functionals to better reproduce known data is not a new one. As already discussed, a great many functional forms have been derived by adjusting parameters semi-empirically to fit known sets of data. Whilst this approach of averaging over a large set of systems and properties can lead to good-quality, widely applicable functionals, there remain some drawbacks.

If the property or system to be investigated differs significantly from the fitting set, then the fitted parameters may no longer be appropriate. Furthermore, the property in question may be too system-dependent, such that any averaged set of parameters cannot produce the desired accuracy. Of particular relevance to this work is that energy linearity has not typically been a primary consideration in the development of conventional functionals, and many problems associated with the delocalisation error still manifest.

A recently proposed solution is to tune the functional parameters to a particular non-empirical condition, on a system-by-system basis, rather than attempting to empirically average over all systems. The justification and implementation of such a procedure is outlined below, and the remainder of the chapter provides a systematic assessment of the technique's application to the delocalisation error.

3.2.1 Tuning to the delocalisation error

Application of the technique to the delocalisation error involves tuning the proportion of exact orbital exchange in global[172,179,180] or range-separated[152,202–214]

hybrid functionals in order to approximately recover Koopmans or other conditions, which are necessary (but not sufficient) for linearity.

The success of this approach lies in the differing behaviour of the HF-like and GGA-like components of a hybrid functional. As is well established, the DFT E_{XC} component causes an incorrect lowering of the energy for fractional N: a convex E *vs* N curve. Conversely, Hartree–Fock (and hence the exact-exchange component of a hybrid functional) incorrectly raises the energy, creating a concave E *vs* N curve and a *localisation* error.

The delicate balance of these two components hence has a cancelling effect, with the overall error being reduced—and so the goal of achieving near-linearity in E *vs* N may be achieved by finding the optimal balance of components. This, of course, is not the whole story. For the functional to be quantitatively useful, the integer points—which are described well by traditional, averaged approaches to functional development—must remain a key consideration. In addition, Karolewski *et al.*[210] recently highlighted a number of caveats for the tuning approach, most notably the violation of size-extensivity.

3.2.2 Satisfying Koopmans' theorem

Although conceptually fairly simple, the idea of tuning the parameters of a functional to attempt to linearise E *vs* N is non-trivial. Clearly, it would be impractical to generate a full E *vs* N curve for each combination of parameters in order to determine the optimal functional for a given system, and even if it was, a reliable means of quantifying the degree of linearity would need to be employed.

Both of these problems can be simplified by considering known aspects of the exact E *vs* N curve. Here we reintroduce the concept of the DFT analogue of Koopmans' theorem within the generalised Kohn–Sham formalism, and use it to identify conditions that can be used as both a measure of, and a constraint on, the linearity of E *vs* N.

For a given species, the piecewise-linear nature of exact E *vs* N—combined with Janak's theorem—gives the exact condition (see Section 2.7.1)

$$\varepsilon_{\text{H}} = -I_0 \tag{3.3}$$

on the electron-deficient side of integer N. In order for an approximate functional to satisfy this equality, two criteria must be met: the elimination of the delocalisation error, and the accurate description of the relative energies (*i.e.* the *calculated* ionisation potential) at integer points.

Of course, in the general case, I_0 is unknown, so we cannot explicitly constrain a functional to reproduce this quantity on a system-dependent basis. We *can*, however, try to enforce the condition $\varepsilon_{\text{H}} = -I$, where I is the ionisation potential calculated as the difference between energies at consecutive integer-N points (ΔSCF). In the absence of delocalisation error, $\varepsilon_{\text{H}} + I = 0$, and hence we introduce our first rudimentary measure of the non-linearity of any functional,

$$H = |\varepsilon_{\text{H}} + I| \,. \tag{3.4}$$

Early manifestations[202,215] of the tuning concept focused on enforcing this Koopmans condition, *i.e.* trying to ensure $H = 0$. For a given functional form one or more of the parameters—in this case the range-separation parameter—are varied until H is minimised, and these parameters are chosen as optimal for this system. As a result, the initial E *vs* N slope on the electron-deficient side of the integer is correct, and the hope is that the remainder of the curve becomes near-linear.

Whilst showing promising results, such a scheme constrains only the electron-deficient side of the integer. We identify a similar exact condition on the electron-abundant side of the integer,

$$\varepsilon_{\text{L}} = -A_0 \,, \tag{3.5}$$

with the corresponding linearity measure (using the ΔSCF electron affinity, A)

$$L = |\varepsilon_{\text{L}} + A| \,. \tag{3.6}$$

For notational purposes, we now identify two segments of the E *vs* N curve for a given species, schematically illustrated in Figure 3.2. We define N_0 to be the (integer) number of electrons in the species in question, and consider the addition and removal of an electron. The electron-deficient part of the curve, between $N_0 - 1$ and N_0 electrons, is then denoted the left-hand side (LHS), and the electron-abundant part (N_0 to $N_0 + 1$) the right-hand side (RHS), of the E *vs* N curve.

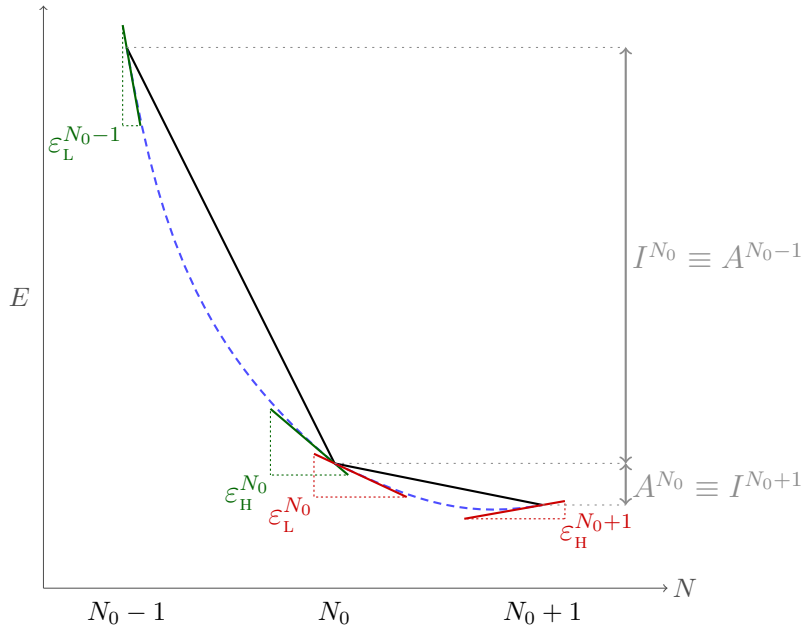

Figure 3.2: Schematic representation of the quantities involved in the tuning methods.

Thus the Koopmans conditions and measures become

$$\varepsilon_{\mathrm{H}}^{N_0} = -I^{N_0}\,, \tag{3.7}$$

$$H_{\mathrm{LHS}} = |\varepsilon_{\mathrm{H}}^{N_0} + I^{N_0}|\,, \tag{3.8}$$

and

$$\varepsilon_{\mathrm{L}}^{N_0} = -A^{N_0}\,, \tag{3.9}$$

$$L_{\mathrm{RHS}} = |\varepsilon_{\mathrm{L}}^{N_0} + A^{N_0}|\,. \tag{3.10}$$

Since these conditions are completely general, we can identify similar conditions on the $(N_0 - 1)$- and $(N_0 + 1)$-electron systems:

$$L_{\mathrm{LHS}} = |\varepsilon_{\mathrm{L}}^{N_0-1} + A^{N_0-1}| \tag{3.11}$$

$$\equiv |\varepsilon_{\mathrm{L}}^{N_0-1} + I^{N_0}|\,, \tag{3.12}$$

and

$$H_{\mathrm{RHS}} = |\varepsilon_{\mathrm{H}}^{N_0+1} + I^{N_0+1}| \tag{3.13}$$

$$\equiv |\varepsilon_{\mathrm{H}}^{N_0+1} + A^{N_0}|\,, \tag{3.14}$$

where we have made use of the fact that $A^{N_0-1} \equiv I^{N_0}$ and $I^{N_0+1} \equiv A^{N_0}$.

Finally, we note that, moving along a given (exact) *E vs N* segment, the relevant frontier orbital remains constant. This introduces the condition that the LUMO at the electron-deficient end of an *E vs N* segment should equal the HOMO at the electron-abundant end. Hence we can introduce two final measures, which we denote Ω,

$$\Omega_{\text{LHS}} = \left|\varepsilon_{\text{L}}^{N_0-1} - \varepsilon_{\text{H}}^{N_0}\right|, \tag{3.15}$$

$$\Omega_{\text{RHS}} = \left|\varepsilon_{\text{L}}^{N_0} - \varepsilon_{\text{H}}^{N_0+1}\right|. \tag{3.16}$$

Each of the quantities H, L and Ω—whilst acting as a measure of the degree of linearity—can also be employed in a tuning procedure. By varying the parameters of a functional (typically either the proportion of exact exchange in a conventional hybrid or the range-separation parameter in an RSH functional) until one of these *tuning norms* is minimised, one can achieve a degree of near-linearity in the relevant *E vs N* segment.

3.2.3 MULTIPLE SEGMENTS

The tuning norms introduced so far each constrain a functional to satisfy a single linearity constraint on one *E vs N* segment of a given system. Arguably, however, this has limited real-world use, where orbital *differences* are often required for meaningful calculations. Take, for example, the fundamental gap, $I_0 - A_0$. As noted by Baer and co-workers,[203,209] in order to accurately estimate this as the difference of Kohn–Sham orbital energies, *both* segments of the *E vs N* curve need to be accurately described.

To achieve this, they introduced new "double-segment" tuning norms, which in our notation are written

$$J_{\text{H, L}} = H_{\text{LHS}} + L_{\text{RHS}} \tag{3.17}$$

and

$$J_{\text{H, H}} = H_{\text{LHS}} + H_{\text{RHS}}, \tag{3.18}$$

where the first and second subscripts indicate the left- and right-hand side tuning conditions, respectively. By analogy, we can also define

$$J_{\text{L, H}} = H_{\text{LHS}} + H_{\text{RHS}} \tag{3.19}$$

and

$$J_{\mathrm{L,\,L}} = L_{\mathrm{LHS}} + L_{\mathrm{RHS}}\,. \tag{3.20}$$

An alternative form of the tuning norm in equation (3.18) was also proposed,[203,209] which we write as

$$\|J_{\mathrm{H,\,H}}\|_2 = \sqrt{H_{\mathrm{LHS}}^2 + H_{\mathrm{RHS}}^2}\,. \tag{3.21}$$

Though conceptually very similar to equation (3.18), a subtle difference arises when the quantity is squared: whereas equation (3.18) contains cross terms, equation (3.21) does not. As before, we may consider the related combinations $\|J_{\mathrm{H,\,L}}\|_2$, $\|J_{\mathrm{L,\,H}}\|_2$ and $\|J_{\mathrm{L,\,L}}\|_2$.

We now observe that equations (3.18) and (3.21) are the first two terms in a series of p-norms,

$$\|J_{\mathrm{H,\,H}}\|_p = \sqrt[p]{H_{\mathrm{LHS}}^p + H_{\mathrm{RHS}}^p}\,, \tag{3.22}$$

which leads us to consider the general p-norm

$$\|J_{x,\,y}\|_p = \sqrt[p]{x^p + y^p}\,, \tag{3.23}$$

where x and y refer to the LHS and RHS conditions respectively; in this work we consider up to $p = 4$. Note that the components of the p-norms are always positive, due to the modulus in the definitions of H, L, and Ω.

Overall this leads to a total of 22 tuning norms, summarised in Table 3.1, which will be assessed in detail later in this chapter. Firstly, though, we present a preliminary investigation into the behaviour of representative RSH functionals, to highlight the system-dependence of the range-separation parameter, and to inform our choice of functional form to use in the systematic assessment.

3.3 Preliminary assessments

3.3.1 functional comparison

An initial assessment was carried out to briefly examine the influence of the RSH functional form on relevant properties, for a few representative systems. Three simple systems—each of which bind an additional electron—were chosen (C, Cu^+ and OH). Each of these was modelled with several common RSH functionals available in Gaussian 09:[216] CAM-B3LYP, LC-ωPBE[118,139,179] and

Table 3.1: Summary of the 22 tuning norms assessed in this chapter.

Tuning norm	Definition
H_{LHS}	$\lvert\varepsilon_{\text{H}}^{N_0} + I^{N_0}\rvert$
L_{LHS}	$\lvert\varepsilon_{\text{L}}^{N_0-1} + A^{N_0-1}\rvert$
Ω_{LHS}	$\lvert\varepsilon_{\text{L}}^{N_0-1} - \varepsilon_{\text{H}}^{N_0}\rvert$
H_{RHS}	$\lvert\varepsilon_{\text{H}}^{N_0+1} + I^{N_0+1}\rvert$
L_{RHS}	$\lvert\varepsilon_{\text{L}}^{N_0} + A^{N_0}\rvert$
Ω_{RHS}	$\lvert\varepsilon_{\text{L}}^{N_0} - \varepsilon_{\text{H}}^{N_0+1}\rvert$
$\lVert J_{\text{H,H}}\rVert_1$	$H_{\text{LHS}} + H_{\text{RHS}}$
$\lVert J_{\text{H,H}}\rVert_2$	$\sqrt{H_{\text{LHS}}^2 + H_{\text{RHS}}^2}$
$\lVert J_{\text{H,H}}\rVert_3$	$\sqrt[3]{H_{\text{LHS}}^3 + H_{\text{RHS}}^3}$
$\lVert J_{\text{H,H}}\rVert_4$	$\sqrt[4]{H_{\text{LHS}}^4 + H_{\text{RHS}}^4}$
$\lVert J_{\text{H,L}}\rVert_1$	$H_{\text{LHS}} + L_{\text{RHS}}$
$\lVert J_{\text{H,L}}\rVert_2$	$\sqrt{H_{\text{LHS}}^2 + L_{\text{RHS}}^2}$
$\lVert J_{\text{H,L}}\rVert_3$	$\sqrt[3]{H_{\text{LHS}}^3 + L_{\text{RHS}}^3}$
$\lVert J_{\text{H,L}}\rVert_4$	$\sqrt[4]{H_{\text{LHS}}^4 + L_{\text{RHS}}^4}$
$\lVert J_{\text{L,H}}\rVert_1$	$L_{\text{LHS}} + H_{\text{RHS}}$
$\lVert J_{\text{L,H}}\rVert_2$	$\sqrt{L_{\text{LHS}}^2 + H_{\text{RHS}}^2}$
$\lVert J_{\text{L,H}}\rVert_3$	$\sqrt[3]{L_{\text{LHS}}^3 + H_{\text{RHS}}^3}$
$\lVert J_{\text{L,H}}\rVert_4$	$\sqrt[4]{L_{\text{LHS}}^4 + H_{\text{RHS}}^4}$
$\lVert J_{\text{L,L}}\rVert_1$	$L_{\text{LHS}} + L_{\text{RHS}}$
$\lVert J_{\text{L,L}}\rVert_2$	$\sqrt{L_{\text{LHS}}^2 + L_{\text{RHS}}^2}$
$\lVert J_{\text{L,L}}\rVert_3$	$\sqrt[3]{L_{\text{LHS}}^3 + L_{\text{RHS}}^3}$
$\lVert J_{\text{L,L}}\rVert_4$	$\sqrt[4]{L_{\text{LHS}}^4 + L_{\text{RHS}}^4}$

ωB97,[217,218] along with a modified parameterisation of CAM-B3LYP, termed LC-B3LYP, where we choose $\alpha = 0$ and $\beta = 1$. Whilst not the optimal parameterisation for many computed properties, this form has the advantage of satisfying the condition $\alpha + \beta = 1$, which is necessary to exhibit the correct asymptotic behaviour of the exchange–correlation potential.

Such a limited set of test data cannot hope to provide firm conclusions, however a number of useful observations can be made. Figure 3.3 compares the variation of the representative norms $\lVert J_{\text{H,H}}\rVert_1$ and $\lVert J_{\text{H,L}}\rVert_1$ with μ for each of the functionals. With the exception of CAM-B3LYP, the functionals show very similar trends, with J reaching a fairly well-defined minimum at a finite value of μ. With CAM-B3LYP, however, J decays asymptotically towards infinite μ, never truly reaching a minimum. This observation lends great weight to the assertion that satisfying $\alpha + \beta = 1$ is a necessary condition for the success of

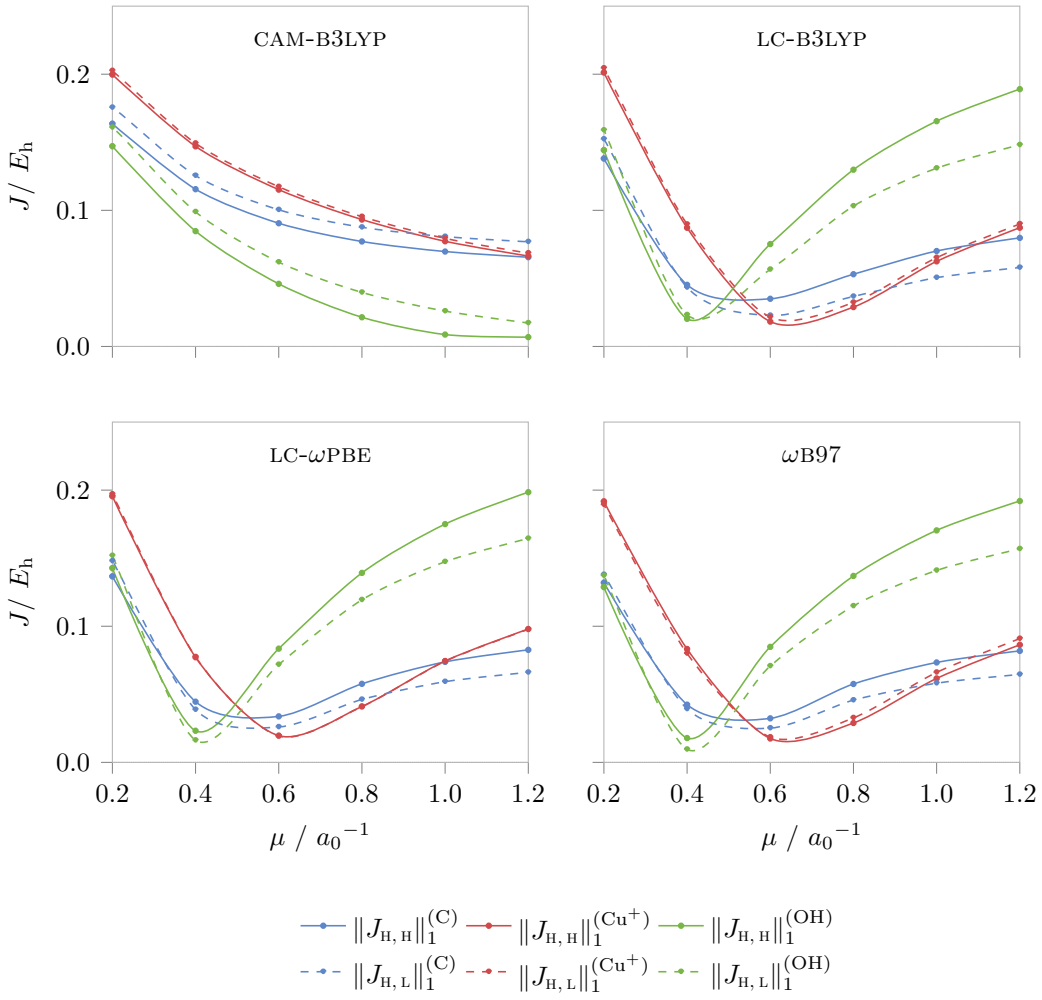

Figure 3.3: Variation of $\|J_{\mathrm{H,H}}\|_1$ and $\|J_{\mathrm{H,L}}\|_1$ with μ for various RSH functional forms, using the aug-cc-pVTZ basis set. The optimal values of μ for each system correspond to minima in the curves (although the curves have been artificially smoothed to guide the eye, the coarse grid of μ values is sufficient to show the qualitative variation of J with μ).

these tuning methods, since CAM-B3LYP is the only functional (of those tested) that does not satisfy this condition.

Another immediately apparent observation is that, aside from CAM-B3LYP, J minimises at similar values for each functional. The precise J_{min}, however, is different for each of the three species, reinforcing the notion that μ needs to be a system-dependent parameter.

Figure 3.4 plots the deviation in the calculated (ΔSCF) I^{N_0} and A^{N_0} from the experimental values[219] $I_0^{N_0}$ and $A_0^{N_0}$, as a function of μ, for each functional. Importantly, the trends vary between functionals much more than the trends in J, and the optimal μ for I^{N_0} or A^{N_0} (*i.e.* that which gives a deviation of $\Delta E = 0$) can vary significantly, both from that of other functionals and from that which minimises J. This has important implications both in the tuning procedure, and in the assessment of the quality of the functional. Specifically, it highlights the fact that we must examine not only the ability of the functional to satisfy linearity conditions (governed by its description of fractional N), but also its ability to reproduce experimental I and A values from ΔSCF calculations (governed by its description of integer N).

3.3.2 SUCCESSIVE IONISATIONS OF CARBON

We have established that the optimal value of μ is sensitive to the system under consideration. However, even for a given atom or molecule, a single value of μ may not be universally appropriate—different values may be required to achieve near-linearity in different segments of the E *vs* N curve. In order to illustrate this, we now investigate the effect of tuning μ for each segment of a given species in turn, by considering successive ionisations of carbon. In effect, we consider the entire E *vs* N curve of carbon, from C^{6+} to C^-, by examining separately the segments between each pair of integer-N species.

Section 3.3.1 established a broad similarity between the various functional forms (where $\alpha+\beta = 1$), and so we concentrate on LC-B3LYP as a representative example, again with the aug-cc-pVTZ basis set. The assessment proceeded as follows. Firstly, an optimal μ, denoted μ^*, was determined for each pair of carbon species, by minimising H for the more-reduced species (thereby satisfying the "HOMO" Koopmans condition). Note here that H refers explicitly to equation (3.8) for the more-reduced species—since we are dealing with multiple E *vs* N segments the use of "LHS" and "RHS" would be potentially ambiguous. Calculations were carried out using the Gaussian 09 software at a range of μ values, with the lowest H chosen as optimal. Next, E *vs* N data were generated for all seven pairs of species, at each μ^*, using a modified version of CADPAC.[220] This was achieved by explicitly incrementing the (fractional) occupation of the relevant frontier orbital (HOMO of the more oxidised species, LUMO of the

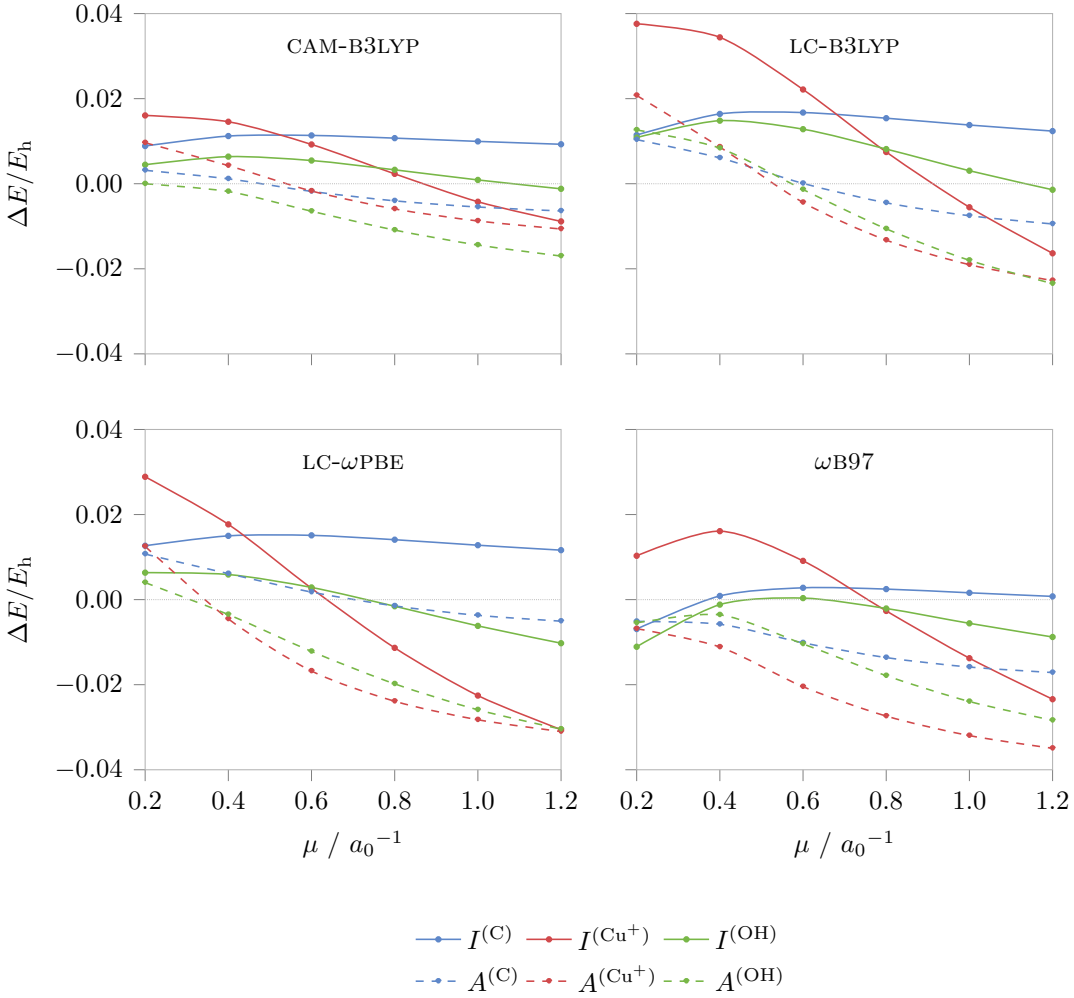

Figure 3.4: Deviation of calculated I^{N_0} and A^{N_0} from experimental $I_0^{N_0}$ and $A_0^{N_0}$, as a function of μ, for various RSH functional forms, using the aug-cc-pVTZ basis set. The optimal values of μ for each system correspond to $\Delta E = 0$.

more reduced) from zero to one in steps of 0.05, and computing the energy at each point.

The procedure of minimising H is illustrated in Figure 3.5, which plots the extent to which the HOMO Koopmans condition is satisfied, as a function of μ, for each carbon species. The optimal μ values—corresponding to where the curves of Figure 3.5 are closest to zero—are summarised in Table 3.2.

It is immediately apparent that there is significant variation in the μ^* for

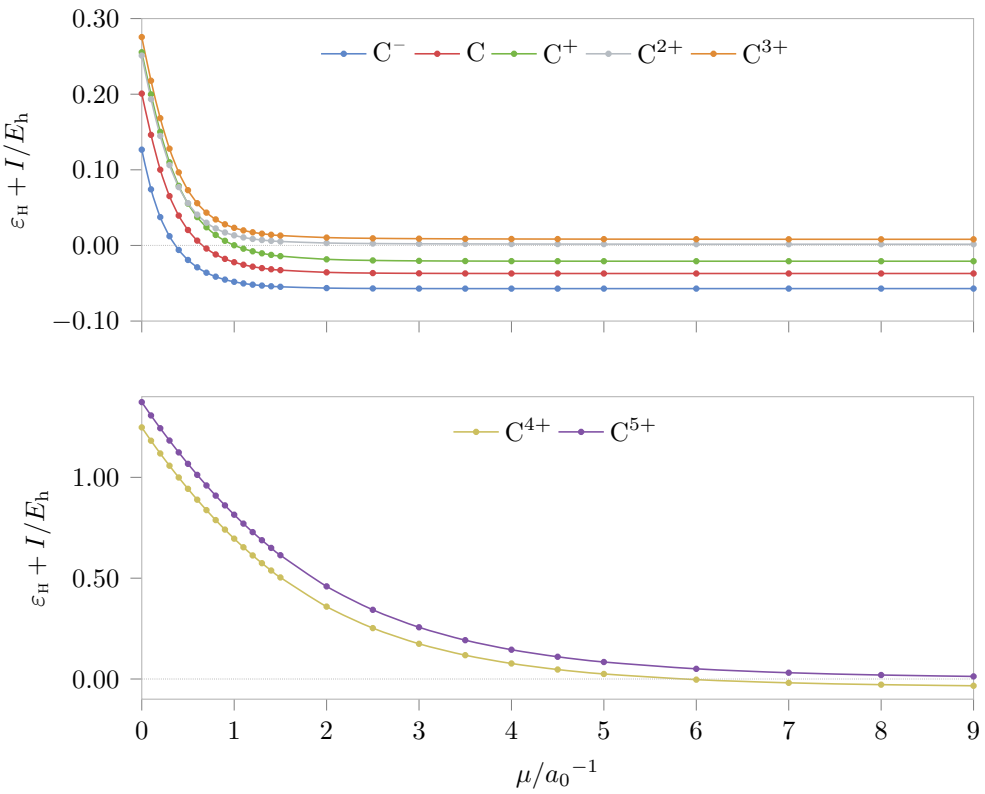

Figure 3.5: The degree to which the HOMO Koopmans condition is satisfied, for successive ionisations of carbon, as a function of μ, using the LC-B3LYP functional and aug-cc-pVTZ basis set. μ^* (Table 3.2) is the value of μ for which H (the absolute values of the plotted data) is minimised.

different E *vs* N segments. The curves for C^-, C, C^+ and C^{4+} each cross $H = 0$ at a finite, well-defined μ^*, and each has a different optimal value. The curves for C^{2+} and C^{5+} merely approach zero as μ tends to infinity, whilst the C^{3+} curve tends asymptotically to a positive value. A particularly significant observation is that μ^* for the C^- and C species differs by $0.3\,a_0{}^{-1}$, and this difference in μ has a big impact on the properties, and hence behaviour, of the optimal functional. Since applying the tuning procedure to these species affects the RHS and LHS E *vs* N segments of neutral carbon, respectively, this immediately presents a challenge for accurately describing both segments, and hence the frontier orbitals of carbon, simultaneously.

Table 3.2: Optimal range-separation parameters μ^* that minimise H for successive ionisations of carbon, using the LC-B3LYP functional and aug-cc-pVTZ.

Species	$\mu^*/{a_0}^{-1}$	
C^-	0.37	
C	0.67	
C^+	1.00	
C^{2+}	∞	$(H \to \sim 0)$
C^{3+}	∞	$(H \to +\text{ve})$
C^{4+}	5.89	
C^{5+}	∞	$(H \to 0)$

E vs N plots

In order to intuitively compare the *E vs N* data for each pair of carbon species on a single plot, and to avoid large differences in scale, we in fact consider what we will denote an *E vs N deviation* curve, plotting ΔE against ΔN. ΔN is defined as the fractional occupancy of the relevant frontier orbital, such that the more oxidised species is described by $\Delta N = 0$, and the more reduced by $\Delta N = 1$. ΔE is the difference between the computed energy and the desired behaviour: an interpolated straight line between the integer-N points. In this way, the closer a plot is to a straight line along zero, the closer it is to being a linear *E vs N* segment.

By construction, the interpolated and calculated energies agree at integer N. For non-integer values, a horizontal line along zero indicates a linear *E vs N* curve, whilst a positive/negative deviation indicates a concave/convex curve.

Figure 3.6 plots this *E vs N* deviation for each of the key μ^* values summarised in Table 3.2 (we choose $\mu = 1000\,{a_0}^{-1}$ as "effectively infinite"). We also present the same variation as computed using Hartree–Fock, for comparison. The plot labels refer to the species at $\Delta N = 1$, *i.e.* that for which H was optimised in determining μ^*. In this way the colours and labels are consistent with Figure 3.5.

It is important to note that the tuning procedure—in this case—only attempts to constrain the initial slope approaching $\Delta N = 1$, and is clearly successful in doing so. For C^-, C, C^+ and C^{4+}, where a finite μ^* is well defined, this behaviour is clearly seen when computed using the appropriate μ. For C^{2+} and C^{5+}, where H decays to zero towards infinite μ, the behaviour is

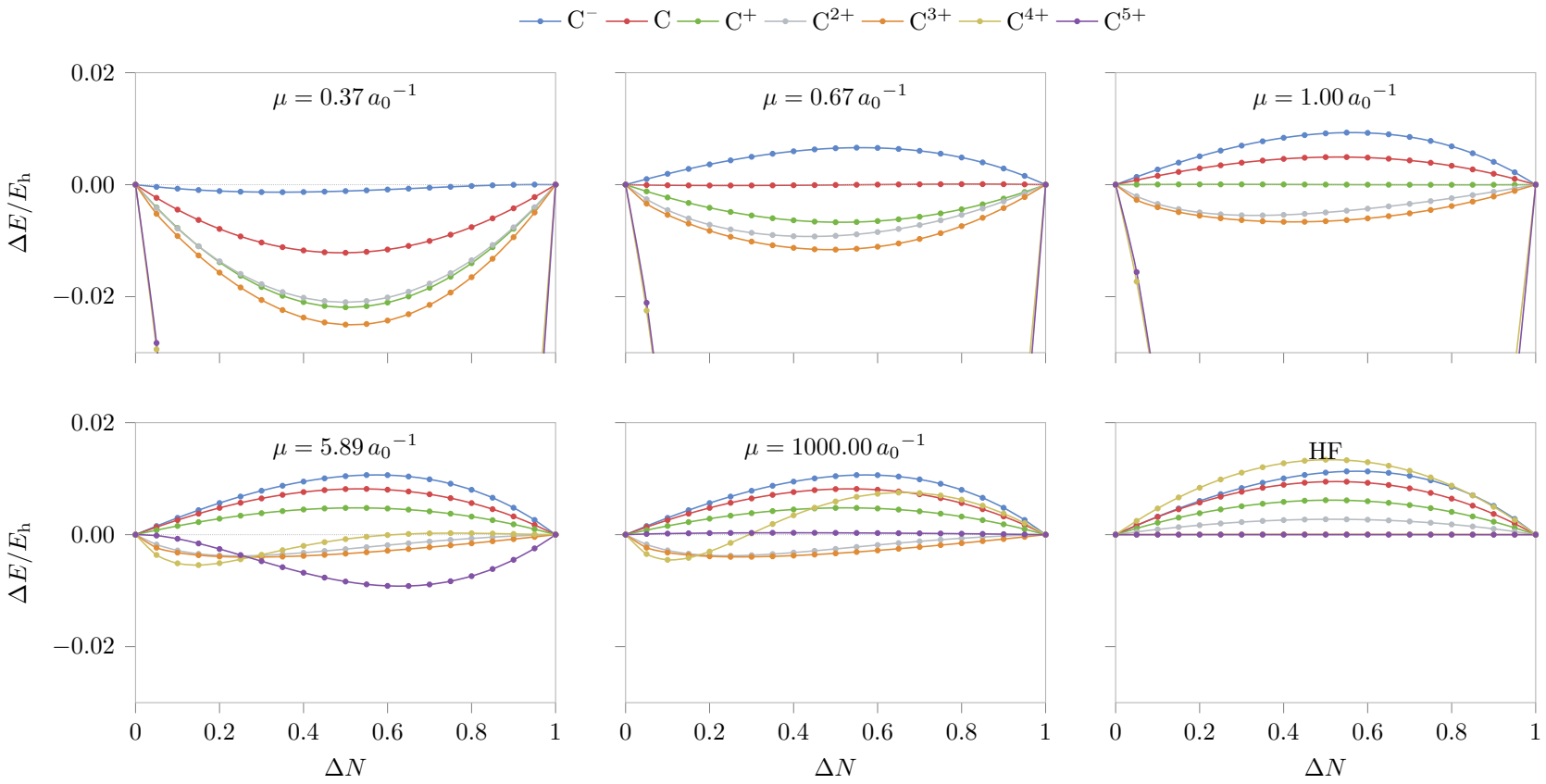

Figure 3.6: *E vs N* deviation curves for successive pairs of carbon species, computed using LC-B3LYP at each of the μ^* values in Table 3.2 and aug-cc-pVTZ, together with Hartree–Fock. Legend labels refer to the species corresponding to $\Delta N = 1$.

approached when computed using very large μ. Finally for C^{3+}, where H decays to a positive asymptote, the behaviour is not seen at all.

The behaviour of the unconstrained end—at $\Delta N = 0$—is slightly less intuitive. If the functional were exact, the slope here would be equal to that at $\Delta N = 1$, and remain constant between $0 \leqslant \Delta N \leqslant 1$. However, this is not the case for any fixed value of μ, and $\varepsilon_{\mathrm{L}}^{\Delta N=0} \neq \varepsilon_{\mathrm{H}}^{\Delta N=1}$. The difference is largely negligible, so the variation between integers is typically very close to linear at μ^*, however with some of the more highly charged species the difference is much more pronounced, leading to distinctly "S-shaped" curves.

These plots perfectly illustrate the significant dependence of E *vs* N curves on the precise value of μ, and reinforce previous observations that it is not possible to enforce linearity on two segments of the curve with a single value of μ. As μ increases, the convexity of each curve decreases, passing through near-linearity at μ^*, before becoming concave. Each curve appears to reach a maximum concavity, such that a further increase in μ no longer affects the curve—as a result the spread of deviations from linearity narrows as μ increases.

Neglecting correlation

It is also interesting to consider the effect of correlation on the above trends. The LC-B3LYP functional with $\mu = 1000\, {a_0}^{-1}$ is effectively equivalent to using the Hartree–Fock functional with (dynamic) correlation added—however the corresponding plot in Figure 3.6 is considerably different than that of pure HF. In particular, the curves for C^{2+} and C^{3+} do not change appreciably above $\mu \approx 1\, {a_0}^{-1}$, whereas with HF they become concave and linear respectively. Similarly, the C^{4+} curve remains S-shaped for $\mu = 1000\, {a_0}^{-1}$, compared with the entirely concave curve for HF.

Figure 3.7 re-plots the variation of H with μ, this time neglecting any correlation in the LC-B3LYP functional. Now, there is a well-defined, finite μ^* for the C^{2+} segment (as reflected in the HF E *vs* N plot, where the curve has passed through near-linearity and become concave), whereas the C^{3+} segment asymptotically approaches zero (reflected in the linear HF plot). Indeed, repeating the E *vs* N analysis using $\mu = 1000\, {a_0}^{-1}$ with no correlation produces a plot—Figure 3.8—identical to that of HF.

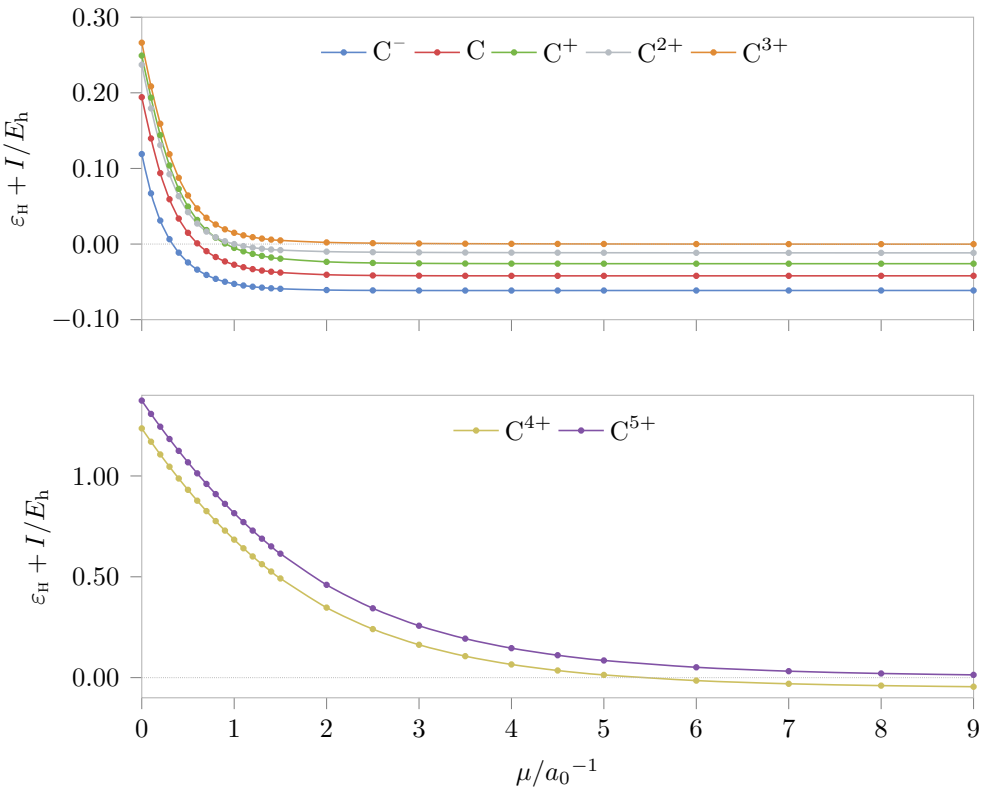

Figure 3.7: The degree to which the HOMO Koopmans condition is satisfied, for successive ionisations of carbon as a function of μ, using correlation-free LC-B3LYP (correlation omitted) and aug-cc-pVTZ.

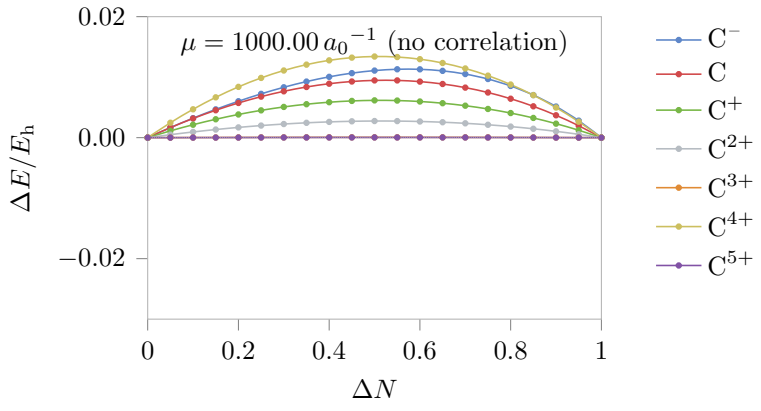

Figure 3.8: *E vs N* deviation curves for successive pairs of carbon species, computed using correlation-free LC-B3LYP at very large μ. Legend labels refer to the species corresponding to $\Delta N = 1$.

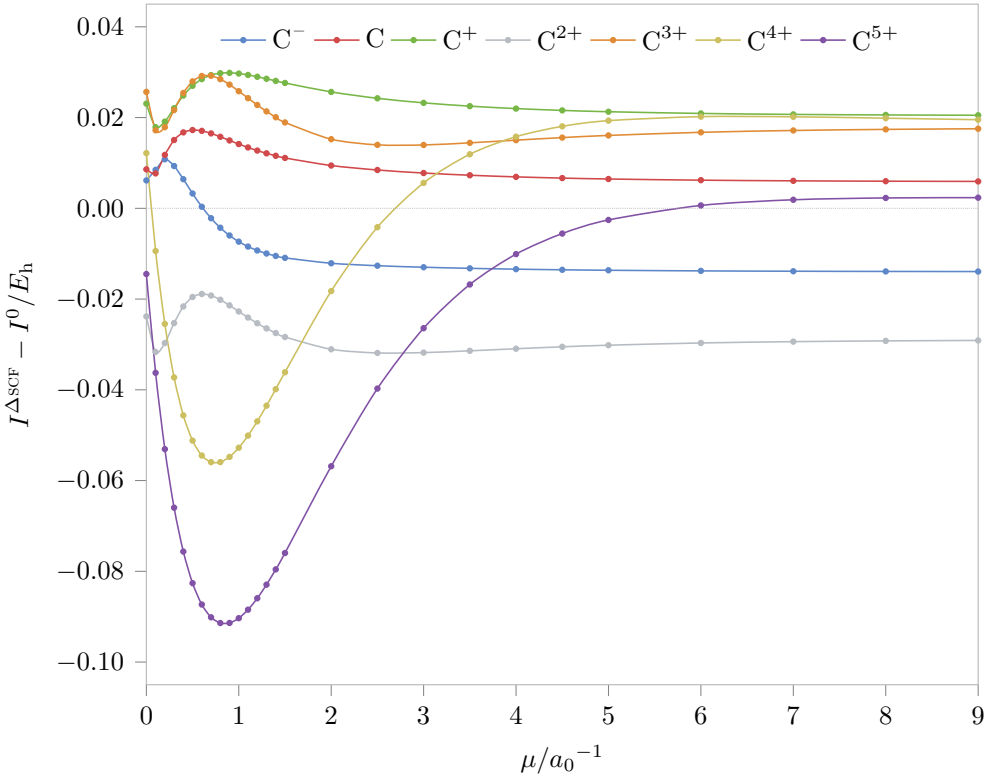

Figure 3.9: Error in calculated (ΔSCF) ionisation energies of carbon species, as a function of μ, for LC-B3LYP.

Comparison to experimental ionisation energy

As discussed in the introduction, we can only explicitly enforce the approximate condition $\varepsilon_\mathrm{H} = -I$, for the obvious reason that in the general case we do not know the exact ionisation energy I_0. However, for the functional to be correct, the condition $I = I_0$ must also be fulfilled, so it is useful—whilst we are dealing with species for which we have exact values—to compare the two quantities.

Figure 3.9 plots the deviation from experiment of successive carbon ionisation energies I, calculated at a range of μ values. It is clear that a broadly similar trend is seen for each removal of the same character of electron—however a different value of μ is optimal for each ionisation energy. Optimal values are listed in Table 3.3. Furthermore, by comparing with Table 3.2, it is clear that the μ that best fulfils $I = I_0$ does *not* necessarily correspond to that which minimises H for the same species. This is consistent with the preliminary assessment of C, Cu^+ and OH.

Table 3.3: Optimal range-separation parameters μ^* that minimise the error in $I^{\Delta\text{SCF}}$ for successive ionisations of carbon, using the LC-B3LYP functional and aug-cc-pVTZ.

Species	$\mu^*/{a_0}^{-1}$
C^-	0.61
C	∞
C^+	0.12
C^{2+}	0.61
C^{3+}	3.0
C^{4+}	0.06 or 2.71
C^{5+}	5.80

In effect we can identify two "optimal" values of μ^* for each carbon species (and hence for each segment of the E *vs* N curve)—one which best minimises the tuning norm (in this case H) and one which best reproduces $I = I_0$. Although the errors in I may still be small when computed at the μ^* appropriate for H, the question remains whether it is more preferable to achieve near-linearity in E *vs* N, or rather to be able to accurately compute I. The carbon analysis has also confirmed that, in general, a different value of μ is required to enforce near-linearity on each E *vs* N segment. How, then, do the double-segment tuning norms behave when attempting to enforce near-linearity on two adjacent segments with a single optimal μ^*? Furthermore, do the μ^* values determined using these norms provide a good estimate of I and A? All of these questions are addressed in the systematic assessment of tuning norms below, and the precise interplay between the various quantities is shown to be central to the success of such methods.

3.4 Assessment of tuning methods

A number of related tuning norms were outlined in Table 3.1, each of which gives a different approach to enforcing known exact conditions on RSH functionals. We now present a systematic study of the performance of each in order to identify which methods are best, and what factors govern their success. We relate this to both the linearity of the E *vs* N dependence, and the errors in various quantities relating to the tuning, namely the frontier orbital energies (and differences), ionisation energies and electron affinities.

3.4.1 Choice of functional

Much of the earlier work on the tuning of RSH functionals, primarily conducted by Baer and co-workers,[202,203,221] used the BNL (Baer–Neuhauser–Livshits) functional form[201,215] to great success. However, as stated in the same works, the techniques are trivially applicable to other RSH functionals. Our preliminary investigations similarly suggested the qualitatively similar behaviour of different RSH forms (where the long-range condition $\alpha + \beta = 1$ was met).

Since an implementation of CAM-B3LYP—trivially modified to LC-B3LYP by changing the α and β parameters—already exists in the modified CADPAC[220] code (which we require to examine the E *vs* N relationships), we choose to proceed with the LC-B3LYP for the full assessment. However, we first provide further justification of its use by reproducing the tuned-BNL results of Refs 203 and 209, and comparing them with equivalent calculations using LC-B3LYP.

The two most prominent tuning norms used in previous works for optimally computing fundamental gaps are (in our notation) $\|J_{\mathrm{H,H}}\|_1$ and $\|J_{\mathrm{H,H}}\|_2$, equations (3.18) and (3.21) respectively. We tuned the functionals by calculating these quantities at a range of μ values for main group atoms from Na to Br (excluding the noble gases, which do not bind an electron). The μ^* for each atom was then identified for each minimised J, and the fundamental gap calculated at this value as $\varepsilon_{\mathrm{L}}^{N_0} - \varepsilon_{\mathrm{H}}^{N_0}$, where N_0 refers to the number of electrons in the neutral atom. We compare this to the experimental fundamental gap, $I_0^{N_0} - A_0^{N_0}$, with data taken from Ref. 219. In the case of nitrogen, which has a negative experimental $A_0^{N_0}$, we explicitly choose the ground-state value of $A_0^{N_0} = 0$ in both the tuning and the analysis.

Figure 3.10 plots the computed fundamental gaps, along with their deviation from experiment, for the BNL functional form implemented in the NWCHEM[222] software (neglecting the subtraction of part of the short-range exchange energy, as noted in Ref. 203), using the aug-cc-pVTZ basis set. Immediately one can see significant differences between the two tuning norms. $\|J_{\mathrm{H,H}}\|_2$ reproduces the values of μ^* published in Refs 209 and 203, with a slight discrepancy in oxygen, with a minor difference of $0.01\,{a_0}^{-1}$ in the μ^* obtained; the effect on the computed fundamental gap is negligible. This difference is largely insignificant, and attributable to slight differences in the code and parameters used (the value of J differs by less than 0.001 eV between the two values of μ).

Of the two methods, $\|J_{\mathrm{H,H}}\|_2$ generally gives a better estimation of the fundamental gap across the range of atoms, with maximum errors of around 0.4 eV. $\|J_{\mathrm{H,H}}\|_1$, on the other hand, does not perform as well overall, despite differing only in the presence of cross-terms when squared. The method gives errors as high as 1 eV, although for certain atoms (*e.g.* phosphorous) it does perform slightly better than $\|J_{\mathrm{H,H}}\|_2$. For nitrogen the two methods are identical, since the RHS criterion is zero. Similarly for the group I and II atoms, the two tuning methods give almost identical results, despite potentially large differences in the μ^* determined by each. The reason for this is how the fundamental gap varies with μ for these atoms: above a certain value, very little variation is seen, so the results are similar regardless of the μ^* obtained.

Figure 3.11 repeats the procedure with LC-B3LYP and the same basis set, using the Gaussian 09[216] software (NB: tests were carried out to ensure consistency between results from the two programs when using an equivalent level of theory). The results closely mimic those obtained using BNL, indicating the suitability of choosing LC-B3LYP for the full systematic assessment.

3.4.2 SYSTEMATIC ASSESSMENT OF TUNING NORMS

Having established that the validity of the approach is independent of the functional form used, we now proceed with a systematic assessment of each of the tuning norms in Table 3.1. We considered the same set of atoms as Baer and co-workers,[204,209] corresponding to those assessed in Figures 3.10 and 3.11, namely Li–F, Na–Cl, and Ga–Br. The ground state spin configuration was used for both neutral atoms and ions.

For each atom, calculations were performed using a series of μ values, and the value that minimised each norm, denoted μ^*, was determined to 2 d.p. For comparison, calculations were also carried out using a representative GGA (PBE), a conventional hybrid functional (B3LYP), and non-tuned range-separated hybrids (BNL and LC-B3LYP, each with range-separation parameter $\mu = 0.4\,a_0{}^{-1}$). Calculations were performed using the Gaussian 09 and CADPAC programs, with the aug-cc-pVTZ basis set.

Inevitably, use of these tuning methods increases the computational effort, since calculations must be performed over a range of μ values. In the present work—where multiple criteria were investigated simultaneously, each with a

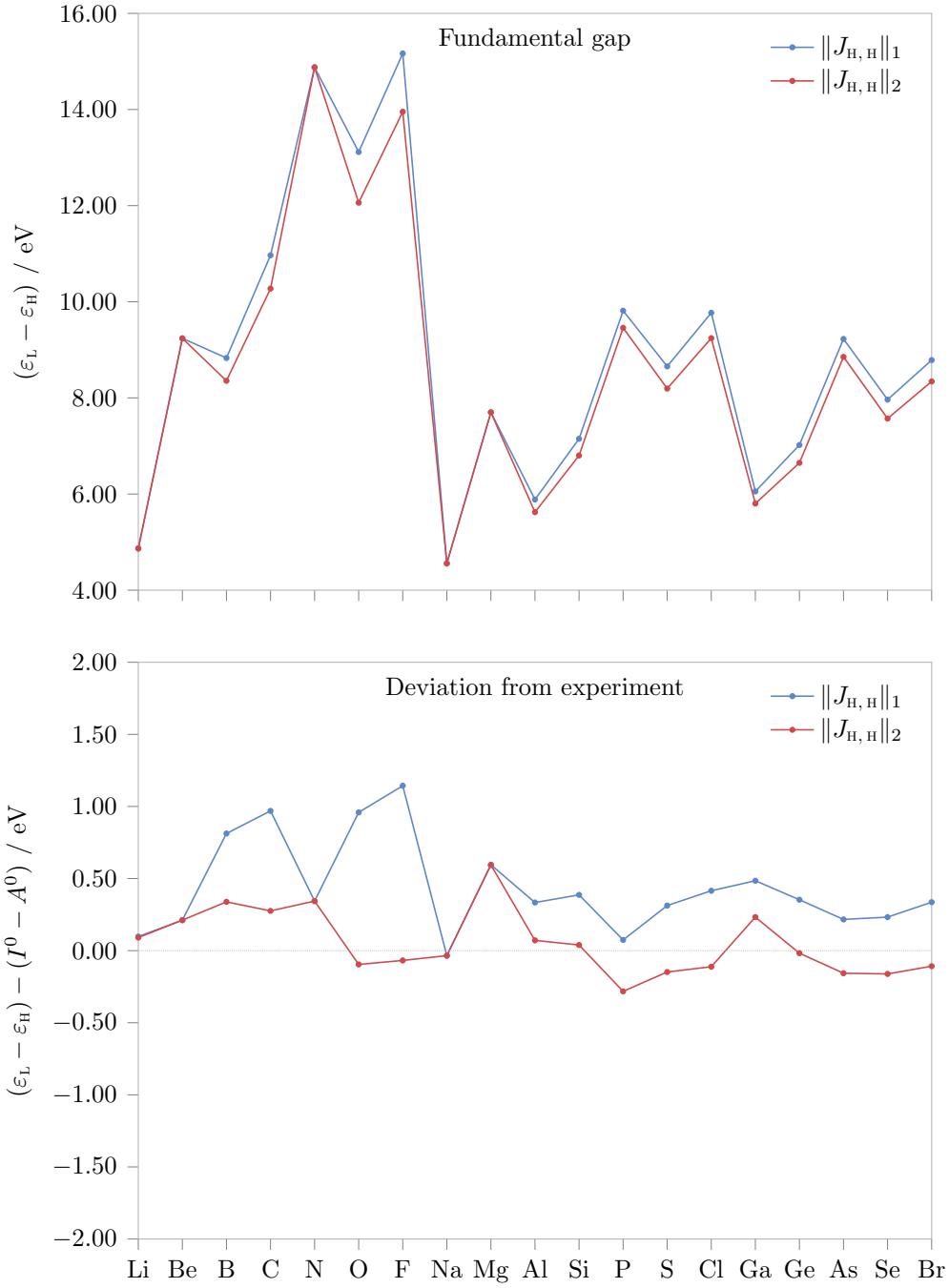

Figure 3.10: HOMO–LUMO gap, and its deviation from experimental fundamental gap, calculated using tuned BNL. All calculations use the aug-cc-pVTZ basis set.

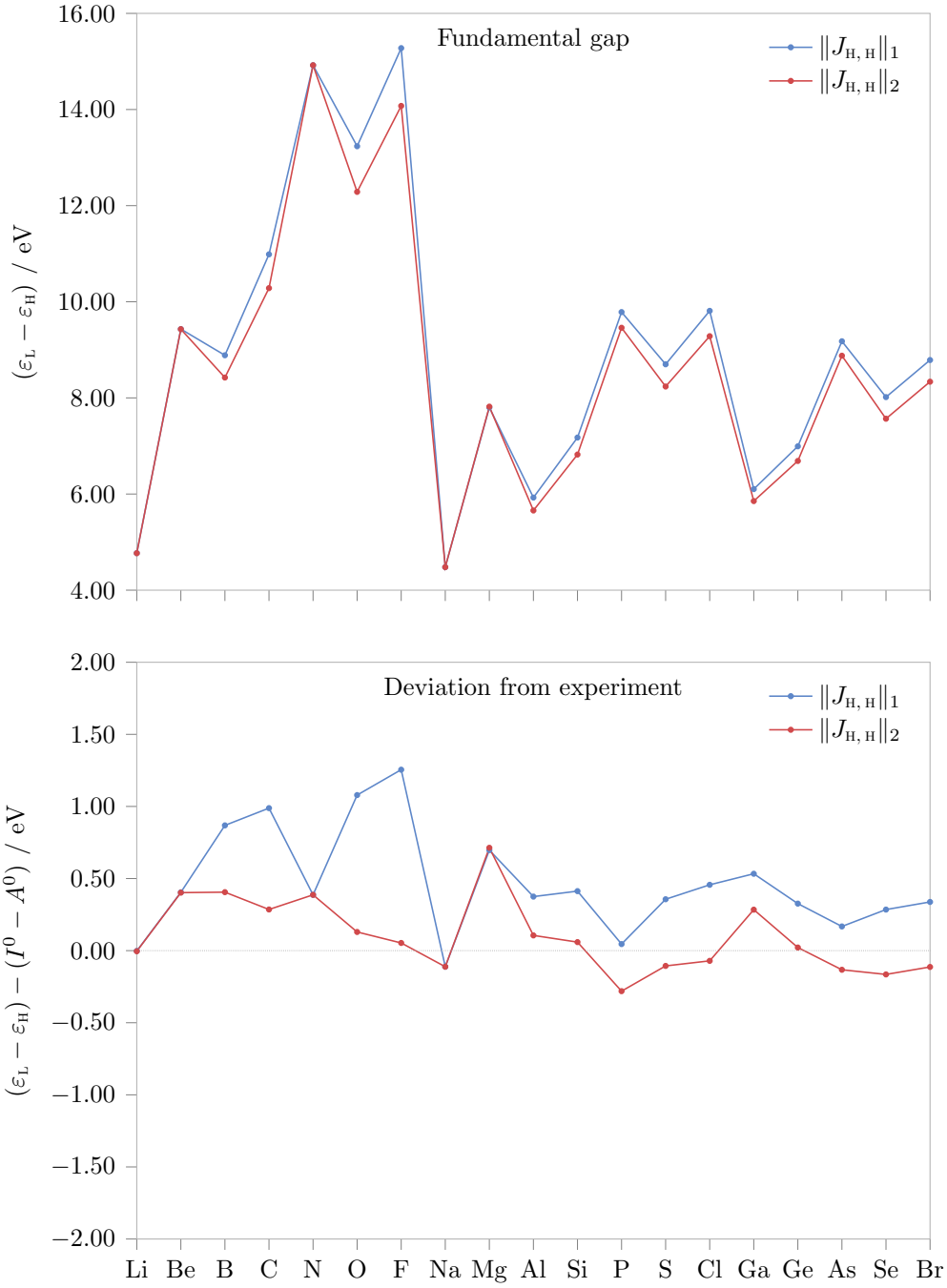

Figure 3.11: HOMO–LUMO gap, and its deviation from experimental fundamental gap, calculated using tuned LC-B3LYP. All calculations use the aug-cc-pVTZ basis set.

(potentially) different μ^*—we used a coarse grid of μ values, refining as necessary close to the minima. For a single tuning norm, the procedure could be carried out more efficiently but would still require a number of separate calculations (typically greater than 20).

We assess the tuning methods in several ways. To test the satisfaction of the Koopmans conditions, we consider the deviation of $\varepsilon_{\mathrm{H}}^{N_0}$ and $\varepsilon_{\mathrm{L}}^{N_0}$ from the exact $-I_0^{N_0}$ and $-A_0^{N_0}$,[223] respectively; it is also pertinent to consider the deviation of the calculated I^{N_0} and A^{N_0} (ΔSCF values determined from integer energy differences) from the exact values. The quantities Ω_{LHS} and Ω_{RHS} provide a measure of linearity for the individual segments, and so we consider the deviation of these quantities from zero (by construction, Ω will be near-zero when successfully employed as the tuning norm, but this will not necessarily be the case for a general tuning norm). Finally, we consider the deviation of $\varepsilon_{\mathrm{L}}^{N_0} - \varepsilon_{\mathrm{H}}^{N_0}$ from the fundamental gap $I_0^{N_0} - A_0^{N_0}$.

Mean absolute deviations are presented in Table 3.4, determined by calculating individual absolute deviations for each atom using its corresponding μ^* value and then averaging over the set of atoms. For clarity, the results are divided into those determined using conventional functionals, and those determined using tuned functionals with single-segment (H, L, and Ω) and double-segment (J) tuning norms.

Conventional Functionals

First, consider the results in Table 3.4 determined using the PBE functional. As expected, the values of $\varepsilon_{\mathrm{H}}^{N_0}$ and $\varepsilon_{\mathrm{L}}^{N_0}$ differ significantly from $-I_0^{N_0}$ and $-A_0^{N_0}$, by 3 eV to 4 eV. By contrast, the directly computed I^{N_0} and A^{N_0} are more than an order of magnitude more accurate. The large values of 6 eV to 8 eV for the linearity measures Ω_{LHS} and Ω_{RHS} quantify the significant E *vs* N curvature. As noted in Section 2.7.1, the errors in $\varepsilon_{\mathrm{H}}^{N_0}$ and $\varepsilon_{\mathrm{L}}^{N_0}$ are of opposite sign and so the deviation is amplified when the difference is used to compute the gap. It follows that the gaps $\varepsilon_{\mathrm{L}}^{N_0} - \varepsilon_{\mathrm{H}}^{N_0}$ deviate by more than 7 eV from $I_0^{N_0} - A_0^{N_0}$. In the case of a GGA such as PBE, the difference between the fundamental gap and the associated orbital energy gap is approximately equal to the exact exchange–correlation integer discontinuity;[149,162] the relationship between curvature and integer discontinuity has recently been discussed by

Table 3.4: Mean absolute deviations (in eV), from exact reference values, of various key quantities computed using conventional functionals and each tuning norm.

	$\varepsilon_{\mathrm{H}}^{N_0}$	$\varepsilon_{\mathrm{L}}^{N_0}$	I^{N_0}	A^{N_0}	Ω_{LHS}	Ω_{RHS}	$\varepsilon_{\mathrm{L}}^{N_0} - \varepsilon_{\mathrm{H}}^{N_0}$
Conventional functionals							
PBE	4.10	3.04	0.15	0.19	8.22	5.71	7.13
B3LYP	3.16	2.35	0.17	0.17	6.57	4.37	5.50
BNL	0.76	0.33	0.15	0.32	1.46	0.50	0.85
LC-B3LYP	0.36	0.37	0.26	0.24	1.29	0.31	0.62
Single-segment tuning norms							
H_{LHS}	0.26	0.33	0.26	0.17	0.16	0.79	0.53
L_{LHS}	0.34	0.40	0.25	0.16	0.17	0.91	0.68
Ω_{LHS}	0.30	0.37	0.26	0.17	0.09	0.85	0.61
H_{RHS}	0.63	0.42	0.26	0.23	1.82	0.21	0.92
L_{RHS}	0.30	0.22	0.26	0.21	1.16	0.21	0.39
Ω_{RHS}	0.46	0.32	0.25	0.23	1.45	0.04	0.65
Double-segment tuning norms							
$\|J_{\mathrm{H,H}}\|_1$	0.24	0.32	0.26	0.17	0.19	0.78	0.51
$\|J_{\mathrm{H,H}}\|_2$	0.09	0.21	0.26	0.19	0.64	0.49	0.19
$\|J_{\mathrm{H,H}}\|_3$	0.10	0.20	0.26	0.19	0.71	0.45	0.17
$\|J_{\mathrm{H,H}}\|_4$	0.11	0.20	0.27	0.19	0.73	0.44	0.16
$\|J_{\mathrm{H,L}}\|_1$	0.25	0.32	0.26	0.17	0.18	0.78	0.52
$\|J_{\mathrm{H,L}}\|_2$	0.17	0.26	0.26	0.18	0.37	0.65	0.34
$\|J_{\mathrm{H,L}}\|_3$	0.16	0.24	0.26	0.18	0.42	0.62	0.31
$\|J_{\mathrm{H,L}}\|_4$	0.14	0.23	0.27	0.18	0.45	0.61	0.28
$\|J_{\mathrm{L,H}}\|_1$	0.33	0.39	0.25	0.16	0.16	0.89	0.67
$\|J_{\mathrm{L,H}}\|_2$	0.11	0.22	0.26	0.19	0.55	0.55	0.22
$\|J_{\mathrm{L,H}}\|_3$	0.11	0.20	0.26	0.19	0.61	0.51	0.19
$\|J_{\mathrm{L,H}}\|_4$	0.11	0.20	0.26	0.19	0.64	0.50	0.17
$\|J_{\mathrm{L,L}}\|_1$	0.34	0.40	0.25	0.16	0.17	0.90	0.67
$\|J_{\mathrm{L,L}}\|_2$	0.22	0.29	0.26	0.18	0.26	0.72	0.44
$\|J_{\mathrm{L,L}}\|_3$	0.19	0.27	0.26	0.18	0.33	0.68	0.38
$\|J_{\mathrm{L,L}}\|_4$	0.18	0.26	0.26	0.18	0.35	0.67	0.37
Mean reference value	−9.78	−1.28	9.78	1.28	0	0	8.50
Reference quantity	$-I_0^{N_0}$	$-A_0^{N_0}$	$I_0^{N_0}$	$A_0^{N_0}$	0	0	$I_0^{N_0} - A_0^{N_0}$

Stein *et al.*[154]

B3LYP performs slightly better than PBE for each of the problematic quantities, reflecting the reduction in convexity caused by the addition of exact exchange, whilst maintaining the low deviations in I^{N_0} and A^{N_0}. The non-tuned range-separated hybrids BNL and LC-B3LYP show a further marked improvement in the problematic quantities, although the deviations remain non-negligible. With these two functionals, the I^{N_0} and A^{N_0} values degrade marginally.

Single-segment tuning

Next consider the functionals tuned to criteria on a single E *vs* N segment, using the H, L, or Ω tuning norms. The deviations in $\varepsilon_{\mathrm{H}}^{N_0}$ and $\varepsilon_{\mathrm{L}}^{N_0}$ are generally close to those from the non-tuned RSH functionals. The lowest deviations in $\varepsilon_{\mathrm{H}}^{N_0}$ and $\varepsilon_{\mathrm{L}}^{N_0}$ are, not surprisingly, obtained by tuning to H_{LHS} and L_{RHS} respectively, since these explicitly optimise the Koopmans conditions that are being assessed. Importantly, these deviations are still non-zero, challenging the assumption that simply tuning to the calculated I^{N_0} or A^{N_0} is sufficient—one must also consider the quality of the calculated I^{N_0} or A^{N_0} itself. In other words, one must take into account the accuracy of the relative energies of the integer-electron systems. In fact, I^{N_0} is in virtually constant deviation by 0.25 eV whichever tuning norm is used, and—somewhat counterintuitively—A^{N_0} is slightly better when tuning to the LHS rather than the RHS.

Tuning to H_{LHS} or L_{LHS} gives comparatively low values of Ω_{LHS}, up to five times smaller than Ω_{RHS}, indicating that the tuning is relatively successful at linearising the LHS segment, but at the expense of the RHS. Tuning to H_{RHS} and L_{RHS} yields analogous behaviour, with Ω_{RHS} values up to nine times smaller than Ω_{LHS}. When used as tuning norms, Ω_{LHS} and Ω_{RHS} give, by construction, near-zero values for their respective linearity measures (although large values are seen on the non-optimised side). However, all that a small value of Ω indicates is that the slopes at the two integers are essentially the same—not necessarily that they are equal to the correct value. In fact, the discrepancy between the deviations in $\varepsilon_{\mathrm{H}}^{N_0}$ and $\varepsilon_{\mathrm{L}}^{N_0}$ and those in the calculated I^{N_0} and A^{N_0} confirms that the slopes are not correct; see Section 3.4.3.

For both LHS and RHS tuning norms, the deviations in the gap $\varepsilon_{\mathrm{L}}^{N_0} - \varepsilon_{\mathrm{H}}^{N_0}$ are of a similar magnitude to the non-tuned RSH functionals. The lowest deviations

are obtained when tuning to H_{LHS} and L_{RHS}, reflecting the fact that these two conditions individually yield the most accurate $\varepsilon_{\text{H}}^{N_0}$ and $\varepsilon_{\text{L}}^{N_0}$, respectively, which are the two components of the gap.

Double-segment tuning

Next consider the functionals tuned to criteria on both E *vs* N segments, using the various J in equation (3.23) as tuning norms. In all cases, the deviations in $\varepsilon_{\text{H}}^{N_0}$ and $\varepsilon_{\text{L}}^{N_0}$ reduce notably from $p = 1$ to $p = 2$, with little subsequent change for $p > 2$. The best results are obtained using the $\|J_{\text{H, H}}\|_p$ series, with deviations as small as 0.1 eV to 0.2 eV. However, similar trends are not seen in the computed I^{N_0} and A^{N_0}—the deviation in I^{N_0} is a near-constant 0.26 eV on average, for all of the tuning norms tested, whereas the deviation in A^{N_0} varies from 0.16 eV to 0.19 eV. This has obvious implications for the tuning methods, which rely on attempting to constrain the frontier orbital energies to these incorrect I^{N_0} and A^{N_0} values.

As p increases beyond unity, Ω_{LHS} increases (implying an *increase* in E *vs* N curvature), which is somewhat counterintuitive given that the error in $\varepsilon_{\text{H}}^{N_0}$ decreases. A more intuitive trend is observed for Ω_{RHS}, which reduces as $\varepsilon_{\text{L}}^{N_0}$ improves. Despite yielding the most accurate $\varepsilon_{\text{H}}^{N_0}$ and $\varepsilon_{\text{L}}^{N_0}$, the values of Ω_{LHS} and Ω_{RHS} are both significant for the $\|J_{\text{H, H}}\|_p$ series with $p \geqslant 2$, indicating that substantial non-linearity remains. Insight into these observations is provided in Section 3.4.3.

The accuracy of the gaps $\varepsilon_{\text{L}}^{N_0} - \varepsilon_{\text{H}}^{N_0}$ again reflect the accuracy of the orbital energies: the best results are obtained for the $\|J_{\text{H, H}}\|_p$ series, with deviations reducing to less than 0.2 eV.

3.4.3 REPRESENTATIVE SYSTEM: THE CARBON ATOM

Further insight into the behaviour of the tuned functionals is obtained by focusing on a single atom, thereby enabling the quantities in Table 3.4 to be explicitly related to E *vs* N plots. We choose the carbon atom, such that $N_0 = 6$, and consider the E *vs* N behaviour in the range $5 \leqslant N \leqslant 7$. Table 3.5 presents μ^* values determined for the carbon atom, using selected tuning norms, along with deviations (calculated minus reference) in $\varepsilon_{\text{H}}^{N_0}$, $\varepsilon_{\text{L}}^{N_0}$, I^{N_0}, A^{N_0}, Ω_{LHS},

Table 3.5: Deviations (in eV), from exact reference values, of various key quantities computed for the carbon atom using selected tuning norms.

	μ^*	$\varepsilon_{\mathrm{H}}^{N_0}$	$\varepsilon_{\mathrm{L}}^{N_0}$	I^{N_0}	A^{N_0}	Ω_{LHS}	Ω_{RHS}	$\varepsilon_{\mathrm{L}}^{N_0} - \varepsilon_{\mathrm{H}}^{N_0}$
Single-segment tuning norms								
H_{LHS}	0.66	−0.47	0.57	0.46	−0.03	−0.05	1.44	1.03
L_{LHS}	0.68	−0.52	0.61	0.45	−0.05	0.05	1.50	1.12
Ω_{LHS}	0.67	−0.49	0.59	0.45	−0.04	0.00	1.47	1.08
H_{RHS}	0.36	0.89	−0.50	0.44	0.21	−2.75	−0.31	−1.38
L_{RHS}	0.43	0.44	−0.14	0.46	0.15	−1.87	0.29	−0.57
Ω_{RHS}	0.39	0.68	−0.33	0.45	0.18	−2.35	−0.03	−1.01
Double-segment tuning norms								
$\|J_{\mathrm{H,H}}\|_1$	0.65	−0.44	0.55	0.46	−0.03	−0.11	1.41	0.99
$\|J_{\mathrm{H,H}}\|_2$	0.53	−0.04	0.24	0.47	0.06	−0.91	0.91	0.29
$\|J_{\mathrm{H,H}}\|_3$	0.52	0.00	0.21	0.47	0.07	−1.00	0.86	0.21
$\|J_{\mathrm{H,H}}\|_4$	0.51	0.04	0.18	0.47	0.08	−1.08	0.81	0.14
Reference value		−11.26	−1.26	11.26	1.26	0	0	10.00
Reference quantity		$-I_0^{N_0}$	$-A_0^{N_0}$	$I_0^{N_0}$	$A_0^{N_0}$	0	0	$I_0^{N_0} - A_0^{N_0}$

Ω_{RHS}, and $\varepsilon_{\mathrm{L}}^{N_0} - \varepsilon_{\mathrm{H}}^{N_0}$, all computed using μ^*. The dependence of these quantities on the choice of tuning norm largely follows the behaviour of the average quantities presented in Table 3.4.

For each tuning norm in Table 3.5, the *E vs N* behaviour was analysed by fixing μ at the corresponding value of μ^* and smoothly varying the number of α-spin electrons in the system from 3 to 5, at fractional intervals (note that due to the similarity between the different double-segment norms, we only present the $\|J_{\mathrm{H,H}}\|_p$ series). The number of β-spin electrons remained fixed at 2. To most effectively illustrate the non-linearity of the *E vs N* curve, we again plot the *deviation* of the calculated energy from a linear interpolation at integer N (*E vs N* deviation curves). As before, the interpolated and calculated energies agree at integer N (by construction), a horizontal line along zero indicates a linear *E vs N* curve, and a positive/negative deviation indicates a concave/convex curve. Results are presented in Figures 3.13 to 3.15.

A pair of straight lines is superimposed onto each *E vs N* deviation curve, one on the electron-deficient side and one on the electron-abundant side of $N = 6$. These lines are the differences between the *exact* piecewise curve and the linear interpolation between calculated integers, aligned at $N = 6$. The

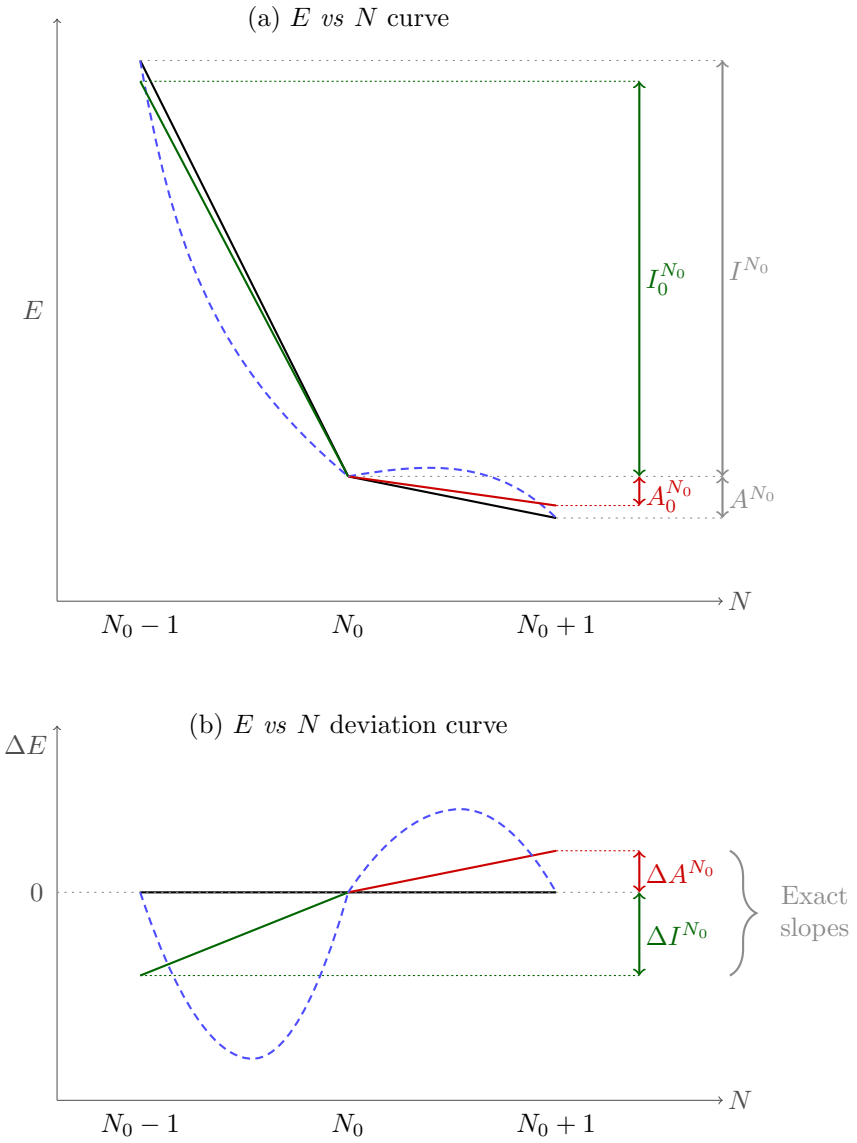

Figure 3.12: Schematic relationship between an E *vs* N curve and an E *vs* N deviation curve. Vertical scale has been exaggerated for clarity.

slopes of these lines indicate the limiting slopes that an E *vs* N deviation curve would have to exhibit at $N = 6$, in order to yield $\varepsilon_{\text{H}}^{N_0} = -I_0^{N_0}$ and $\varepsilon_{\text{L}} = -A_0^{N_0}$, respectively. We term them "exact slopes", in the context of an E *vs* N deviation plot. Figure 3.12 illustrates the relationship between a conventional E *vs* N curve, with exact I and A values superimposed, and an E *vs* N *deviation* curve, with *exact slopes* superimposed.

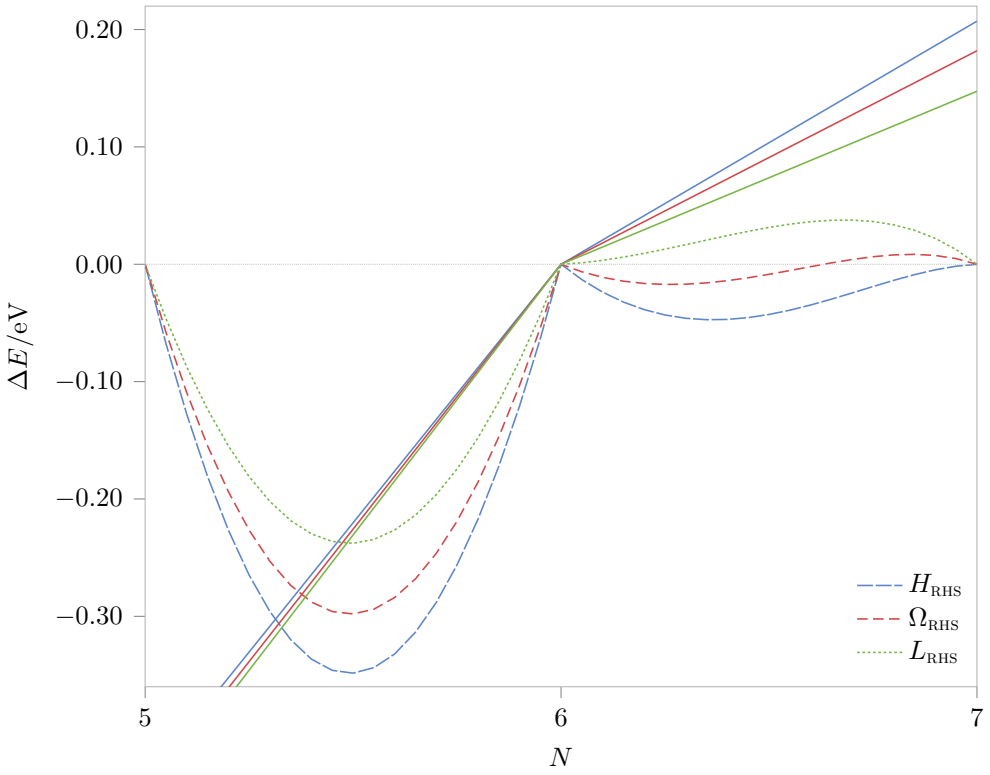

Figure 3.13: E *vs* N deviation curves (dashed/dotted curves) and exact slopes (solid straight lines) for the carbon atom using RHS tuning norms.

These exact slopes provide a useful guide to the quality of a functional: the difference between the slope of the E *vs* N deviation curve and the exact slopes quantifies the deviations in $\varepsilon_{\text{H}}^{N_0}$ and $\varepsilon_{\text{L}}^{N_0}$ from $-I_0^{N_0}$ and $-A_0^{N_0}$, whilst the deviation of the exact slope from *horizontal* quantifies the deviations in I^{N_0} and A^{N_0}, again from $I_0^{N_0}$ and $A_0^{N_0}$. We note that if the exact slope is not horizontal then satisfaction of the exact Koopmans conditions ($\varepsilon_{\text{H}}^{N_0} = -I_0^{N_0}$; $\varepsilon_{\text{L}} = -A_0^{N_0}$) will *require* non-linearity in the E *vs* N curve.

Figure 3.13 presents the E *vs* N deviation curves for the three μ^* values determined by tuning to the RHS. The three μ^* values are rather different from one another and so the E *vs* N behaviour of each is also quite different. In moving from H_{RHS} to Ω_{RHS} to L_{RHS}, the slopes on either side of $N = 6$ move closer to the exact slopes, and so the deviations in $\varepsilon_{\text{H}}^{N_0}$ and $\varepsilon_{\text{L}}^{N_0}$ in Table 3.5 reduce. The exact slopes are notably offset from horizontal, reflecting the deviations in I^{N_0} and A^{N_0} in Table 3.5. By construction, tuning to H_{RHS} and

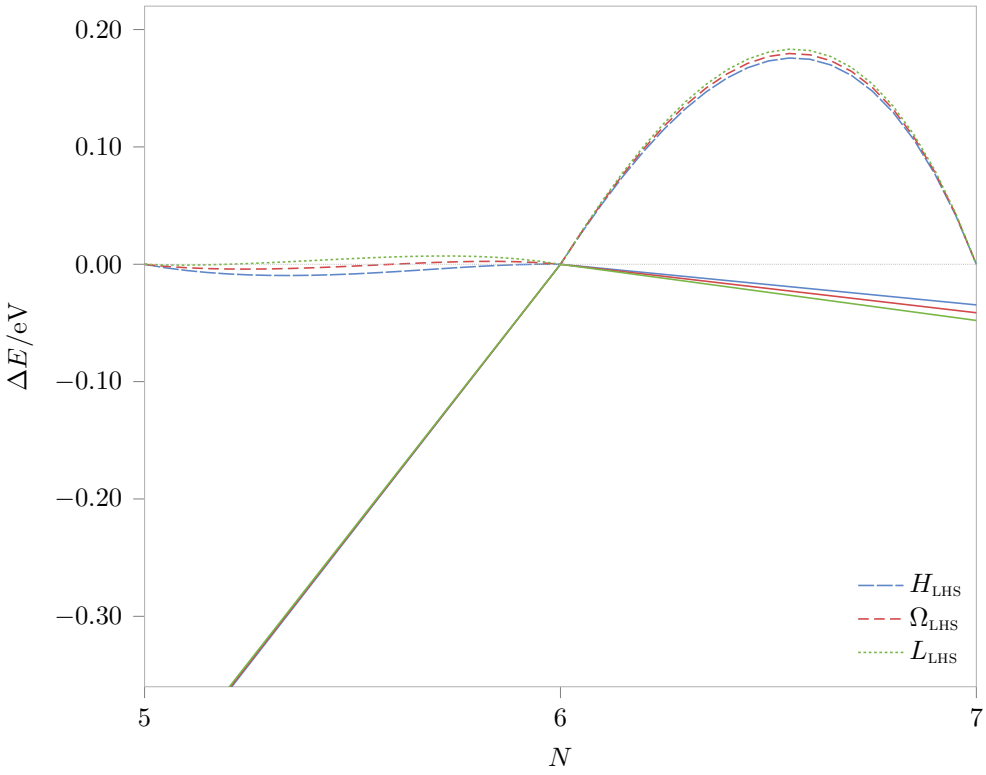

Figure 3.14: *E vs N* deviation curves (dashed/dotted curves) and exact slopes (solid straight lines) for the carbon atom using LHS tuning norms.

L_{RHS} yields—for the RHS segment—near-zero slopes at $N = 7$ and $N = 6$ respectively. However, in both cases the unconstrained end of the RHS segment exhibits a much larger slope, leading to the small but non-negligible Ω_{RHS} values in Table 3.5; by contrast the LHS segment is highly convex, with large Ω_{LHS} values. By construction, tuning to Ω_{RHS} yields essentially identical slopes at $N = 6$ and $N = 7$ for the RHS segment and hence near-zero values for Ω_{RHS}, but the slopes themselves are not zero, resulting in a curve with a point of inflection. The LHS segment is again highly convex, with a correspondingly large Ω_{LHS} value.

Figure 3.14 shows the analogous curves when tuning to the LHS. The three μ^* values are now very close to one another and so the differences in the *E vs N* behaviour are much less pronounced. The near-linearity of the LHS is much more pronounced than the RHS was in Figure 3.13 and so, at first sight, one might expect an accurate $\varepsilon_{\text{H}}^{N_0}$. However, the plot simply illustrates

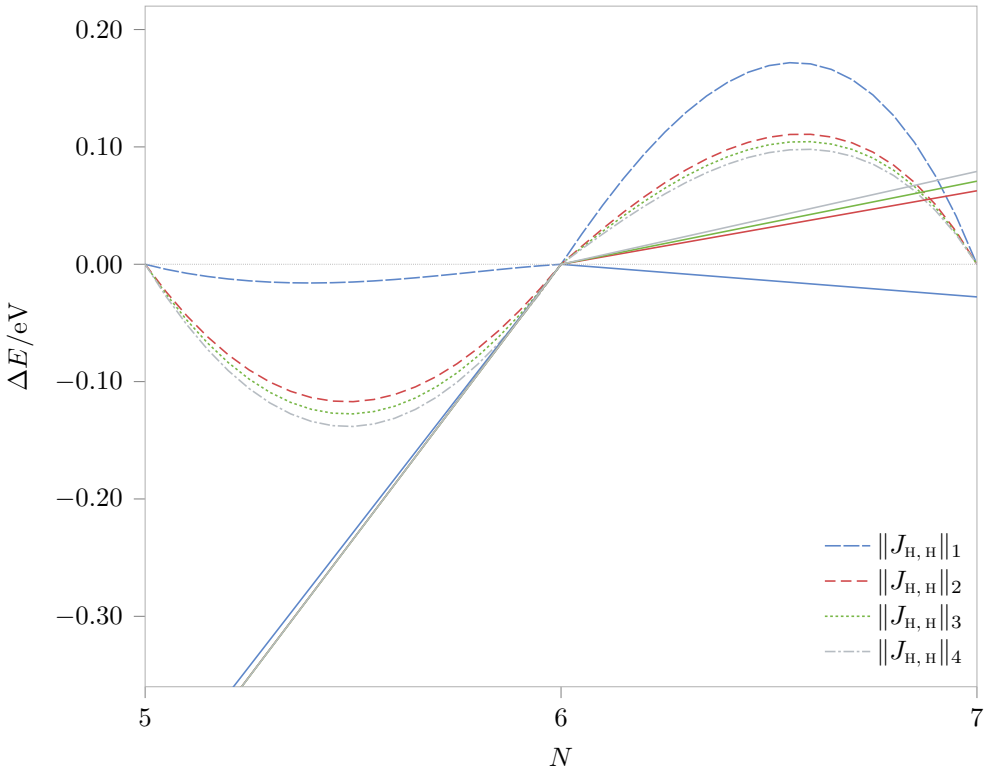

Figure 3.15: *E vs N* deviation curves (dashed/dotted curves) and exact slopes (solid straight lines) for the carbon atom using double-segment tuning norms.

that $\varepsilon_{\mathrm{H}}^{N_0} \approx -I^{N_0}$; by contrast the deviation of the exact slope from horizontal indicates that $I^{N_0} \neq I_0^{N_0}$ and so the discrepancy between $\varepsilon_{\mathrm{H}}^{N_0}$ and $-I_0^{N_0}$ is actually significant. The deviation in $\varepsilon_{\mathrm{L}}^{N_0}$ is of a similar magnitude—in this case, the deviation arises largely due to the curvature, rather than the error in A^{N_0}, which is now much smaller.

Finally, consider the curves obtained by tuning to both segments. Each set of p-norms shows a similar trend, so we choose the most successful method, $\|J_{\mathrm{H,H}}\|_p$, to illustrate the behaviour. Figure 3.15 presents the *E vs N* deviation curves using the μ^* values obtained for $1 \leqslant p \leqslant 4$. When $p = 1$, $\mu^* = 0.65\, {a_0}^{-1}$, which is essentially the same as the value obtained when tuning to the LHS segment alone. As a result, the corresponding curve is close to those in Figure 3.14. Increasing p to 2 yields $\mu^* = 0.53\, {a_0}^{-1}$, roughly midway between the LHS-only and RHS-only optimised values. Appropriately, this leads to some reduction in concavity on the RHS (Ω_{RHS} in Table 3.5 decreases) but

increased convexity on the LHS (Ω_{LHS} increases in magnitude), so that neither side shows near-linear behaviour. As p is increased to 3 and 4, μ^* decreases marginally again, with a corresponding small shift in the E *vs* N deviation curves. Despite the lack of linearity on either side, using $p > 1$ yields slopes that are closest to the exact slopes, and hence the values of $\varepsilon_{\text{H}}^{N_0}$ and $\varepsilon_{\text{L}}^{N_0}$ are optimal. The good performance of the $p > 1$ functionals therefore arises from a convenient error cancellation between lack of linearity and errors in I^{N_0} and A^{N_0}.

3.5 Conclusions

Tuning the range-separation parameter of an RSH functional (or, indeed, simply the proportion of exact exchange in a conventional hybrid) in order to better reproduce desired conditions, is a relatively recent concept which is gaining increasing popularity. The ability to obtain good-quality approximations—from existing methods—of problematic quantities previously inaccessible to conventional DFT methods, is an enticing prospect.

Through preliminary investigations involving C, Cu^+ and OH, along with successive ionisations of carbon, we confirmed the system-dependence of the optimal range-separation parameter μ, and its dependence on the quantity being assessed. By comparing various functional forms, we concluded that the long-range condition $\alpha + \beta = 1$ is desirable not only for the asymptotic behaviour of the exchange–correlation potential, but also for the tuning norms to exhibit a well-defined minimum. Despite this, we found that RSH functionals which *do* satisfy this conditional are qualitatively interchangeable when utilised in the tuning procedure.

We performed a systematic assessment of a range of tuning norms for enforcing approximate energy linearity, *i.e.* reducing delocalisation error through a system-by-system optimisation of a representative RSH functional. For a series of atoms, the accuracy of frontier orbital energies, ionisation potentials, electron affinities, and orbital energy gaps was quantified, with particular attention paid to the extent to which approximate linearity was actually achieved in the resulting E *vs* N curve.

The tuning approaches can yield significantly improved orbital energies

and orbital energy gaps, compared to those from conventional functionals. For a given N_0-electron system, optimal results were obtained using a tuning norm, $\|J_{\mathrm{H},\mathrm{H}}\|_{p>1}$, based on the HOMO energy of the N_0- and (N_0+1)-electron systems, with deviations of just 0.1 eV to 0.2 eV in these quantities, compared to exact values. However, detailed examination of the behaviour of the carbon atom illustrates a subtle cancellation of errors. Specifically, the very fact that it is not possible to achieve near-linearity on both the LHS and RHS of N_0, means that at the optimal μ there is an inherent non-linearity remaining in E *vs* N. Conveniently, the degree of curvature—and hence the deviation of the frontier orbital energies from the computed ionisation potentials and electron affinities to which μ was tuned—is of approximately the same magnitude as the errors (compared to experiment) of I and A themselves.

The implications of this observation, and thus the validity of the approach, are debatable. On the one hand, one is taking advantage of a convenient cancellation of errors, and so the robustness of the method is immediately called into question. On the other, regardless of the intricacies of the method, the approach provides a seemingly reliable means of computing quantities beyond the capability of conventional functionals, by a trivial parameter-tweak and without the need to modify or implement any new functional forms.

Whilst we are inclined to agree with the latter argument, we urge that care must be taken in the interpretation of any results, and hope that our observations provide a useful insight into the properties of these tuned functionals, which may assist the future development of functionals free from the delocalisation error.

3.5.1 Further considerations

Despite the apparent success of these tuning methods, there are a number of remaining drawbacks. Firstly, the procedure requires a series of calculations on the N_0-, (N_0-1)- and/or (N_0+1)-electron systems in order to narrow down the optimal μ. For our test set of systems, with the simultaneous investigation of many tuning norms, we initially chose a coarse grid of μ values, optimising with a tighter grid around the minima to achieve the desired precision. Whilst arguably the most efficient compromise between CPU time and man-hours, the number of individual calculations quickly ran into the thousands. This is an

extreme case, and the procedure for a single system/tuning norm is much easier, however even with the implementation of a relatively efficient minimisation routine (complicated by the need to compute I and A for each μ) we estimate several tens of calculations would still be necessary.

The system-dependence of the approach is paramount to its success, yet undeniably remains one of its drawbacks, in that μ must be recomputed for any new system, in the time-intensive manner outlined above. Likewise the optimal μ is unique to the property for which it has been tuned, as determined by the tuning norm used. If calculation of a different property is desired then a different μ may be optimal, requiring a full set of new tuning calculations based on a different norm. This may not be a problem if one simply desires a good estimation of a particular property for one or two molecules, but could quickly become unmanageable for a large set of systems.

Finally, we note the lack of size-extensivity in the method, hindering its use for, say, dissociation energies. Despite its drawbacks, the method is an important step towards reducing or even eliminating the delocalisation error.

In this work we have focused entirely on a range-separation scheme under the constraint that $\alpha = 0$ and $\beta = 1$, with μ the only free parameter. An interesting future investigation could be made into two- and three-parameter optimisations (the former maintaining the desirable constraint that $\alpha + \beta = 1$), in addition to alternative forms of the short- and long-range partitioning. In particular, given that a non-zero α can be beneficial for thermochemical properties,[200] the convenient cancellation of errors between the calculated ionisation energy and any residual non-linearity may be affected.

In the next chapter, we examine the effect of the tuning procedure on the *electron density*, which—like the energy—should exhibit a linear variation with fractional N. By considering the Fukui function (the derivative of the density with respect to N, and a key quantity in conceptual DFT) computed with tuned RSH functionals, we investigate whether μ^* values that are optimal for near-linearity in E *vs* N are similarly optimal for the density.

4

Fukui function from tuned functionals

As with E *vs* N, the variation of the exact density with N is also piecewise-linear, with discontinuities in its derivative at integer N. In this chapter, we consider whether functionals tuned to give near-linearity in E *vs* N also give near-linearity in the density. We do this by considering the influence of the range-separation parameter on the Fukui function in conceptual DFT. We relate our findings to calculations of the Mulliken electronegativity, another important quantity in conceptual DFT, which can be determined from the frontier orbital energies and so should benefit from an improved description of E *vs* N.

4.1 Conceptual DFT

Broadly speaking, the term "conceptual DFT" refers to the field of relating known empirical chemical concepts, such as electronegativity, to specific definitions extracted from DFT quantities—in particular successive derivatives of the energy functional. Whilst a brief outline is provided below, full discussions can be found in chapters 4 and 5 of Parr and Yang,[58] and a comprehensive review by Geerlings *et al.*[53]

4.1.1 Chemical potential and electronegativity

Electronegativity is an important concept for understanding and predicting chemical reactivity, and describes the degree to which an atom in a molecule draws electron density towards itself. Originally proposed by Pauling[225] in terms of a relative scale, multiple alternative definitions now exist.[226–233] The Mulliken expression, in particular, involves only the ionisation potential and electron affinity of the atom and so is independent of any relative scale; for this reason it is often termed the *absolute* electronegativity,[234,235] and is readily identifiable with quantities computed using DFT.[51–53]

Recall the definition of the electronic energy E in terms of the density $\rho(\boldsymbol{r})$,

$$E[\rho] = F[\rho] + \int \rho(\boldsymbol{r}) v(\boldsymbol{r})\, d\boldsymbol{r}\,, \tag{4.1}$$

where $F[\rho]$ is a universal functional of the density and $v(\boldsymbol{r})$ is the external potential. The energy change dE of a system passing from one ground state to another (*i.e.* perturbation of the system due to a chemical reaction) can be written as

$$dE = \int \left(\frac{\delta E}{\delta \rho(\boldsymbol{r})}\right)_v d\rho(\boldsymbol{r})\, d\boldsymbol{r} + \int \left(\frac{\delta E}{\delta v(\boldsymbol{r})}\right)_\rho dv(\boldsymbol{r})\, d\boldsymbol{r}\,, \tag{4.2}$$

From the Euler–Lagrange equation, equation (2.19), we can write the first term of equation (4.2) as

$$dE_v = \int \mu\, d\rho(\boldsymbol{r})\, d\boldsymbol{r} \tag{4.3}$$

$$= \mu\, dN\,, \tag{4.4}$$

where μ is the Lagrange multiplier, or chemical potential (*not* the range–separation parameter tuned in Chapter 3), the v subscript indicates fixed $v(\boldsymbol{r})$, and we have used the quadrature relation for the number of electrons N in equation (2.12).

The second term of equation (4.2) can be identified by considering the functional derivative of equation (4.1) with respect to $v(\boldsymbol{r})$ at fixed ρ, which—since F is a universal functional of ρ—is simply $\rho(\boldsymbol{r})$, hence

$$dE_\rho = \int \rho(\boldsymbol{r})\, dv(\boldsymbol{r})\, d\boldsymbol{r}\,. \tag{4.5}$$

Combining equations (4.4) and (4.5) we can rewrite equation (4.2) as

$$dE = \mu\, dN + \int \rho(\boldsymbol{r})\, dv(\boldsymbol{r})\, d\boldsymbol{r}\,. \tag{4.6}$$

Alternatively, the energy can be considered as a functional of the number of electrons N and $v(\boldsymbol{r})$, perturbations in which give the differential

$$dE = \left(\frac{\partial E}{\partial N}\right)_v dN + \int \left(\frac{\delta E}{\delta v(\boldsymbol{r})}\right)_N dv(\boldsymbol{r})\, d\boldsymbol{r}\,. \tag{4.7}$$

By comparing terms in equations (4.6) and (4.7), we thus identify the individual first-order derivatives of the energy as

$$\mu = \left(\frac{\partial E}{\partial N}\right)_v \tag{4.8}$$

and

$$\rho(\boldsymbol{r}) = \left(\frac{\delta E}{\delta v(\boldsymbol{r})}\right)_N . \tag{4.9}$$

Through Iczkowski and Margrave's definition[230] of electronegativity,

$$\chi = -\left(\frac{\partial E}{\partial N}\right) , \tag{4.10}$$

we can now relate the Lagrange multiplier of the Euler equation—the chemical potential—directly to this well-established chemical concept,

$$\mu = -\chi\,, \tag{4.11}$$

thus cementing the link between computational and conceptual DFT.

It must be noted that, as a result of the derivative discontinuity, $\mu(N)$ is a stepwise function, discontinuous at integer N. This leads to two electronegativity values, χ^- and χ^+, when evaluated on the LHS or RHS of the integer respectively, and corresponding to an electrophilic ($dN < 0$) or nucleophilic ($dN > 0$) perturbation. Typically, a finite difference approach is taken for the calculation of χ^+ and χ^- for a given species with $N = N_0$ electrons:

$$\chi^- = E_{N_0-1} - E_{N_0} \quad = I\,, \tag{4.12}$$

$$\chi^+ = E_{N_0} \quad - E_{N_0+1} = A\,, \tag{4.13}$$

with the electronegativity taken to be the average of the two,

$$\chi = \frac{\chi^+ + \chi^-}{2} = \frac{I + A}{2}\,. \tag{4.14}$$

The latter relationship is equivalent to Mulliken's definition of the electronegativity.

4.1.2 Fukui function and higher derivatives

Having established the link between first derivatives of the energy with core chemical concepts, it is logical to consider higher derivatives. An overview of common functions and their definitions is given in Scheme 4 of Ref. 53. Of particular note are the second derivatives: the electronic Fukui function

$$f(\boldsymbol{r}) = \left(\frac{\partial\, \delta E}{\partial N \delta v(\boldsymbol{r})}\right), \tag{4.15}$$

the chemical hardness

$$\eta = \left(\frac{\partial^2 E}{\partial N^2}\right)_{v(\boldsymbol{r})}, \tag{4.16}$$

and the linear response function

$$\chi(\boldsymbol{r}, \boldsymbol{r}') = \left(\frac{\delta^2 E}{\delta v(\boldsymbol{r}) \delta v(\boldsymbol{r}')}\right)_N. \tag{4.17}$$

The Fukui function is a chemical reactivity index that acts as a generalisation[236,237] of Fukui's FMO (frontier molecular orbital) theory.[238–240] By resolving equation (4.15), it can be identified either as the sensitivity of the chemical potential to a perturbation at point $\boldsymbol{r}$, or as the change in the electron density $\rho(\boldsymbol{r})$, at each point $\boldsymbol{r}$, when the total number of electrons N is varied,

$$f(\boldsymbol{r}) = \left(\frac{\delta \mu}{\delta v(\boldsymbol{r})}\right)_N = \left(\frac{\partial \rho(\boldsymbol{r})}{\partial N}\right)_v. \tag{4.18}$$

From the latter definition it is clear that this quantity is again subject to the derivative discontinuity at integer N, and so for any system with $N = N_0$ electrons, we may consider derivatives on both the electron-abundant and electron-deficient sides,

$$f^+(\boldsymbol{r}) = \left(\frac{\partial \rho(\boldsymbol{r})}{\partial N}\right)^+_{v(\boldsymbol{r})}, \tag{4.19}$$

and

$$f^-(\boldsymbol{r}) = \left(\frac{\partial \rho(\boldsymbol{r})}{\partial N}\right)^-_{v(\boldsymbol{r})}. \tag{4.20}$$

Physically these correspond to a generalisation of Fukui's FMO reactivity indices, measuring the reactivity of a site towards a nucleophilic attack provoking

the addition (f^+) or an electrophilic attack provoking the subtraction (f^-) of an electron. The third index—representing reactivity towards a radical reagent—is given by the average

$$f^0 = \frac{f^+(\boldsymbol{r}) + f^-(\boldsymbol{r})}{2}\,. \tag{4.21}$$

In practice, these quantities are usually approximated by a finite difference method,

$$f^+(\boldsymbol{r}) \approx \rho_{N_0+1}(\boldsymbol{r}) - \rho_{N_0}(\boldsymbol{r})\,, \tag{4.22}$$

and

$$f^-(\boldsymbol{r}) \approx \rho_{N_0}(\boldsymbol{r}) - \rho_{N_0-1}(\boldsymbol{r})\,, \tag{4.23}$$

although an analytic expression has been developed by Yang *et al.*.[241]

4.2 Near-linearity of the density gradient

In Chapter 3, our goal was to tune the parameters of a given functional until the electronic energy E varied linearly (or, in reality, almost linearly) with the number of electrons N, between each integer pair of N. The electron density should also exhibit similar linear variation, and this has obvious implications for the Fukui function, which is defined above as the change in ρ as N is varied.

We may generalise the finite difference (FD) methods to

$$f^+(\boldsymbol{r}) \approx \frac{\rho_{N_0+x}(\boldsymbol{r}) - \rho_{N_0+y}(\boldsymbol{r})}{x-y}\,, \tag{4.24}$$

and

$$f^-(\boldsymbol{r}) \approx \frac{\rho_{N_0-y}(\boldsymbol{r}) - \rho_{N_0-x}(\boldsymbol{r})}{x-y}\,, \tag{4.25}$$

where $0 \leqslant y < x \leqslant 1$. The "usual" definitions given by equations (4.22) and (4.23) use values of $x = 1$ and $y = 0$, corresponding to the densities at integer numbers of electrons. For the exact functional—*i.e.* where the linearity condition holds—the value of $f^{\pm}(\boldsymbol{r})$ should be identical for any combination of x and y. Intuitively though, we would expect this *not* to be the case for

traditional density functional approximations that suffer from the delocalisation error.

We can illustrate this relatively simply by plotting the difference between the integer FD functions, equations (4.22) and (4.23), and alternatives calculated using equations (4.24) and (4.25) with a small fractional N, *i.e.* choosing $y = 0$ and $x = 0.01$. Pictorially, this is analogous to comparing the ΔSCF energy differences I and A with the limiting initial slopes ε_{H} and ε_{L}, for an E *vs* N plot.

These *Fukui difference plots* were produced as follows:

$$\Delta f^{+} = f^{+}_{x=1} - f^{+}_{x=0.01} \tag{4.26}$$

$$= (\rho_{N_0+1} - \rho_{N_0}) - \frac{(\rho_{N_0+0.01} - \rho_{N_0})}{0.01}\,, \tag{4.27}$$

and

$$\Delta f^{-} = f^{-}_{x=0.01} - f^{-}_{x=1} \tag{4.28}$$

$$= \frac{(\rho_{N_0} - \rho_{N_0-0.01})}{0.01} - (\rho_{N_0} - \rho_{N_0-1})\,, \tag{4.29}$$

where we have dropped the $(\boldsymbol{r})$ for clarity. A straight line along zero is the desired behaviour, indicating linearity in the (initial) density gradient.

An example is shown in Figure 4.1 for the carbon atom, using BLYP as a representative GGA, and the aug-cc-pVTZ basis set. Since we are no longer dealing with a one-dimensional quantity, we choose to plot along a single coordinate, in this case the z axis; such a choice is arbitrary as long as the direction is consistent across all calculations. The density was thus computed along a grid of z values, for a carbon atom with 5, 6, 7, 5.99 and 6.01 electrons, taking care to add or remove electrons (or fractions thereof) to and from the same orbitals; the atom was oriented such that the p_z orbital was the HOMO of the cationic species (and so was always occupied).

It can immediately be seen that for both Δf^{+} and Δf^{-} there are notable deviations from zero, indicating a discrepancy between the initial density gradient, between 0 and 0.01 electrons, and the overall density change between 0 and 1 electrons. The discrepancy is particularly pronounced for Δf^{-}. Naturally, we would expect this behaviour to improve, as before, when moving to hybrid and RSH functionals, however it is particularly interesting to consider the effect

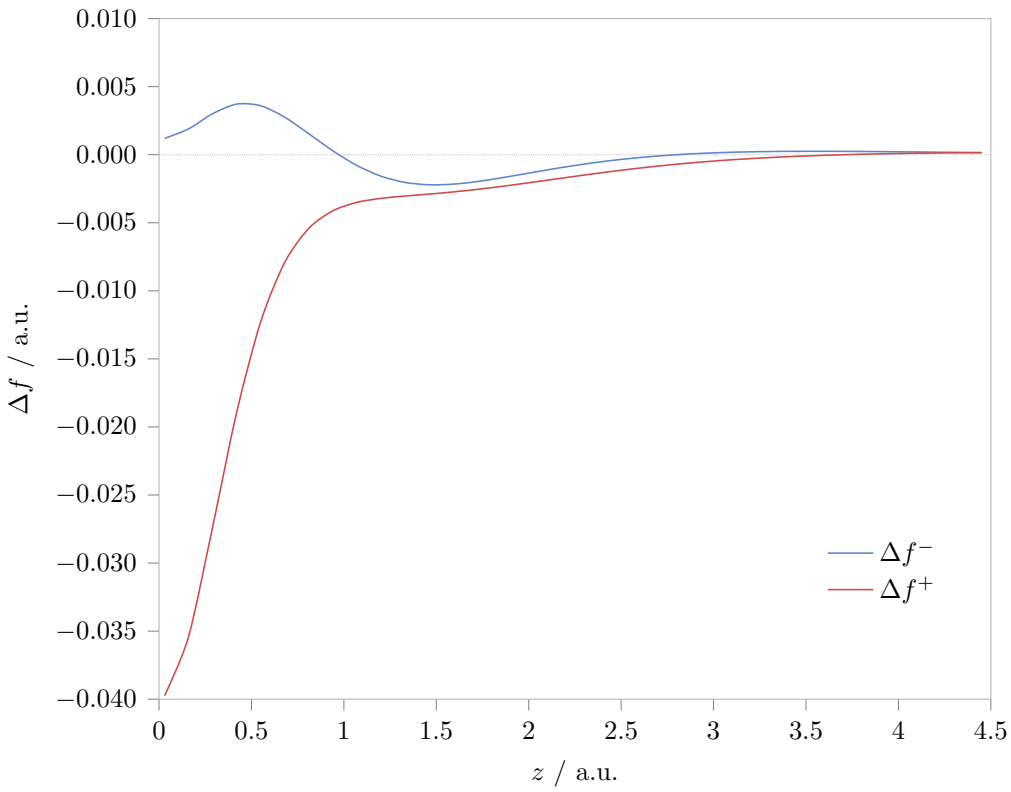

Figure 4.1: Example Fukui difference plots for the carbon atom, using BLYP, showing the difference between finite difference Fukui functions computed using $N_0 \pm 1$ and $N_0 \pm 0.01$ electrons. All calculations use the aug-cc-pVTZ basis set.

of the RSH tuning procedure on the Fukui difference plots. Specifically, are the optimal range-separation parameters determined in Chapter 3 from norms appropriate for *energy* linearity similarly appropriate for the *density* variation, in the context of the Fukui function?

4.3 Functional tuning

One could follow a similar procedure to that in Chapter 3: determine a tuning norm—based now on desired criteria for the *density*, and minimise it for a given RSH functional to determine the optimal μ^* (note that here we revert to using μ to denote the range-separation parameter; for the remainder of the chapter we will discuss the chemical potential in terms of the electronegativity

χ to avoid confusion). However, finding suitable minimisation criteria for the density—which is a function of $\boldsymbol{r}$—is not as trivial as for the energy. Instead, we consider whether functionals tuned to give near-linearity in E *vs* N are similarly optimal for the density, by examining Fukui difference plots computed at a range of μ values (including μ^* values determined from energy tuning norms).

We first examine, separately, the Fukui functions for carbon, computed using the finite difference method with $x = \pm 1$ and $x = \pm 0.01$, respectively. Calculations using BLYP and CAM-B3LYP are compared with those using LC-B3LYP with a selection of μ values. The values of $\mu = 0.66\,{a_0}^{-1}$ and $\mu = 0.43\,{a_0}^{-1}$ correspond to tuning the H_{LHS} and L_{RHS} norms from Chapter 3, *i.e.* enforcing near-linearity in the LHS and RHS E *vs* N segments, respectively. The value of $\mu = 0.4\,{a_0}^{-1}$ corresponds to a "typical" choice, for an LC functional with $\alpha = 0$ and $\beta = 1$, when averaged for thermochemical properties. Finally, the value of $\mu = 0.35\,{a_0}^{-1}$ investigates the effect of further reducing the proportion of short-range exact exchange. After presenting the Fukui functions themselves, we examine the Fukui *difference* plots, to investigate the extent of linearity in the density variation with respect to N.

4.3.1 RESULTS

Figure 4.2 plots the Fukui functions for carbon, computed using the usual finite difference method from integer-electron systems, *i.e.* equations (4.22) and (4.23). It is clear that there is very little difference between the six functionals. This is fully consistent with the ability of each of the functionals to model integer-electron systems relatively successfully, and with a similar degree of accuracy to one another.

Figure 4.3 plots the *fractional* Fukui functions for carbon, this time calculated with equations (4.24) and (4.25) using $x = \pm 0.01$, for the same set of functionals, on the same vertical scale. Now, significant differences can be seen between the functionals, reflecting their respective abilities to model systems with fractional numbers of electrons. Immediately, this would suggest that the Fukui difference plots will exhibit a significant functional-dependence. The biggest variation between functionals occurs in the core regions, below around $2\,a_0$; as z increases, the differences between functionals become much

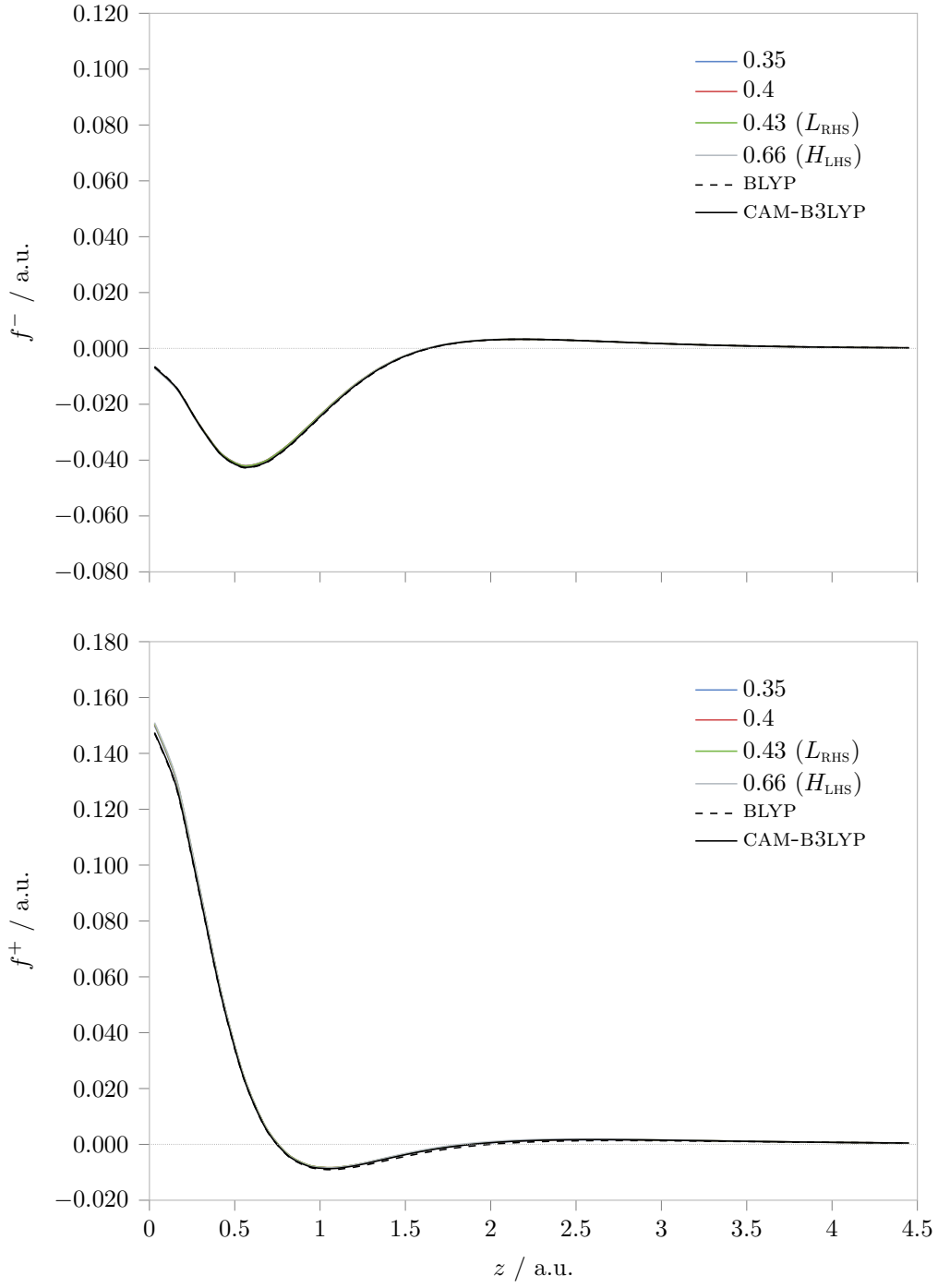

Figure 4.2: Fukui functions for the carbon atom, computed using the finite difference method from integer-electron species, equations (4.22) and (4.23).

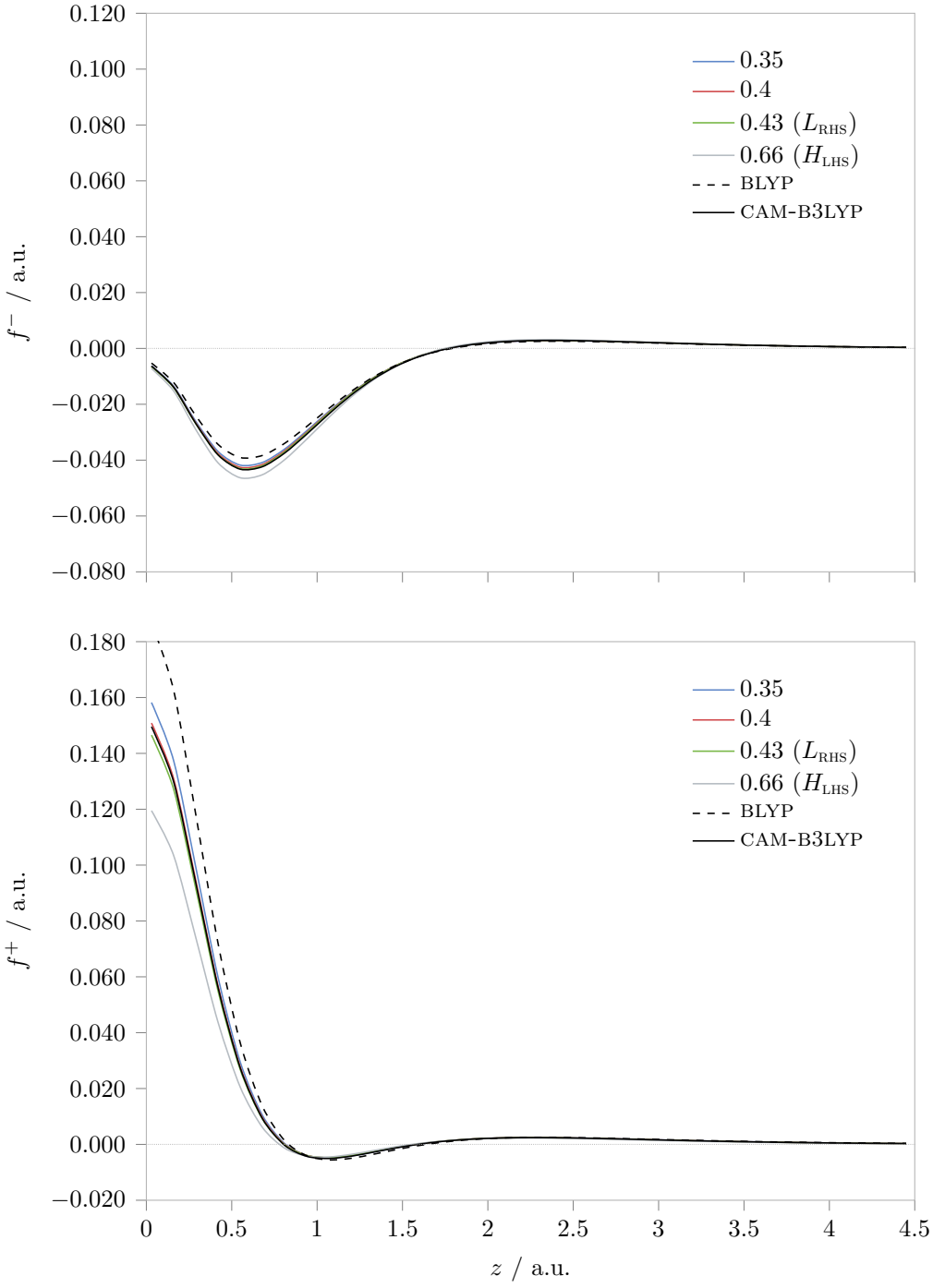

Figure 4.3: Fractional Fukui functions for the carbon atom, computed using the generalised finite difference method, equations (4.24) and (4.25), with $x = \pm 0.01$.

less pronounced.

Figure 4.4 shows the Fukui difference plots for the carbon atom. It is immediately evident that the presence of exact exchange has a significant impact on the shape of the plot. The most prominent illustration of this can be seen by comparing BLYP, a GGA, with CAM-B3LYP, a conventional RSH functional. The difference plots for BLYP show significant deviation from zero, whereas those for CAM-B3LYP—whilst still not linear—show a marked improvement over much of space, highlighting the beneficial effect of the incorporated exact exchange.

When varying the value of μ in LC-B3LYP, it would be reasonable to hypothesise that the LHS-tuned value (in terms of E *vs* N) of $0.66\,{a_0}^{-1}$ might give the optimal Δf^-, whilst the RHS-tuned value of $0.43\,{a_0}^{-1}$ would give the optimal Δf^+. However, it in fact transpires that $0.43\,{a_0}^{-1}$ appears better in *both* cases, whereas $0.66\,{a_0}^{-1}$ produces a deviation similar in magnitude, though opposite in sign, to the GGA. Furthermore, a value of $0.4\,{a_0}^{-1}$—deemed appropriate for most thermochemical properties—is slightly better still (as, in fact, is the thermochemically averaged CAM-B3LYP). Reducing μ to $0.35\,{a_0}^{-1}$ shifts the plots back in the direction of the GGA, suggesting that insufficient exact exchange is included.

The biggest variation in Δf^- and Δf^+ between functionals occurs in the core region, below around $z = 2\,a_0$. The chemical relevance of the Fukui function, however, is particularly associated with the region of space corresponding to the van der Waals radius: approximately $3.2\,a_0$ (1.7 Å) for carbon. Figure 4.5 shows an expansion of the Fukui difference plots in this region of space. In general, all of the functionals perform well in this region: only very small deviations from zero are seen in the difference plots.

On the electron-deficient side, BLYP in fact shows the smallest deviation from zero, whilst for the RSH functionals, the deviation again decreases with decreasing μ. The value of $\mu = 0.66\,{a_0}^{-1}$ optimised for *energy* linearity on the electron-deficient side, is again least optimal.

On the electron-abundant side, the ordering of functionals is the same, however the plots are shifted downwards. It is again evident that for most values of z, decreasing μ from $0.66\,{a_0}^{-1}$ to $0.4\,{a_0}^{-1}$ is beneficial.

Overall, it is clear that despite only small variations in the chemically

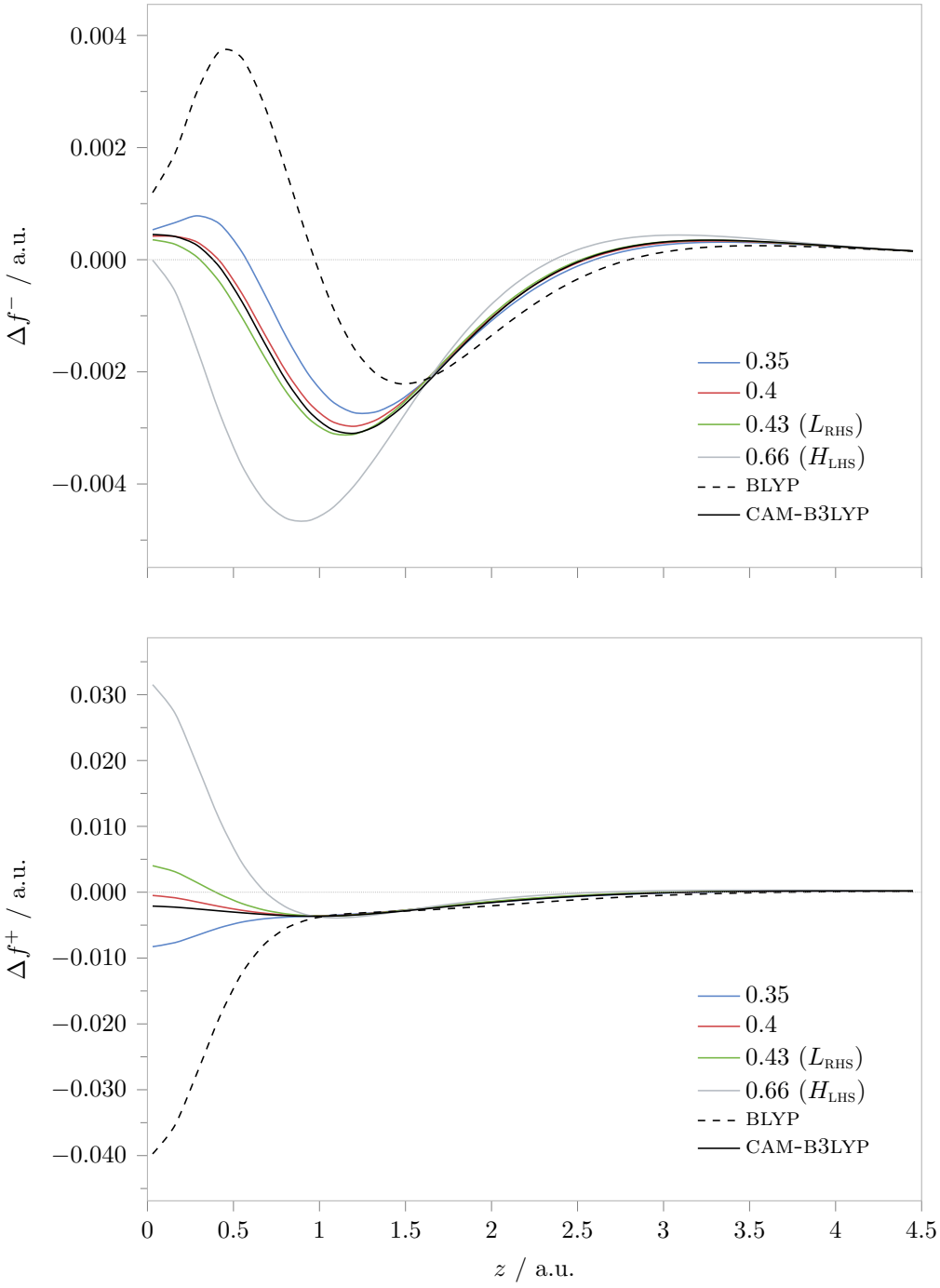

Figure 4.4: Fukui difference plots for the carbon atom, computed with LC-B3LYP at a variety of μ values, compared to representative GGA and RSH functionals. All calculations use the aug-cc-pVTZ basis set.

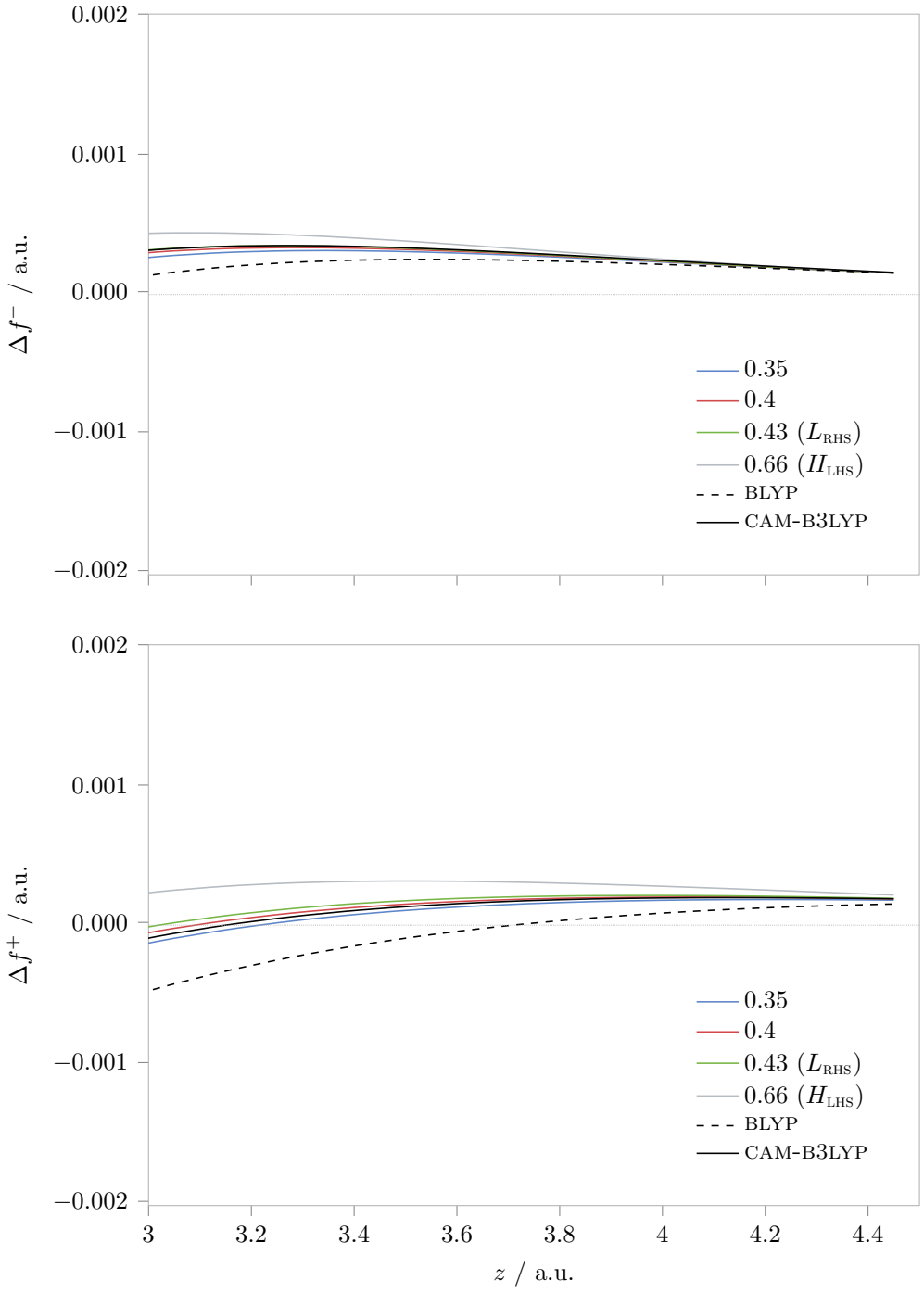

Figure 4.5: Expansion of the chemically significant region of the carbon Fukui difference plots, around the van der Waals radius. All calculations use the aug-cc-pVTZ basis set.

Table 4.1: Errors in the electronegativity χ from the reference value of 6.26 eV, calculated using experimental I and A values.[219] Errors in χ calculated using I and A, equation (4.14), are compared with those calculated using ε_{H} and ε_{L}, equation (4.30).

Functional	Equation (4.14)	Equation (4.30)
BLYP	0.12	−0.63
CAM-B3LYP	0.21	−0.30
0.35	0.26	−0.20
0.4	0.32	−0.17
0.43	0.31	−0.15
0.57	0.25	−0.09
0.66	0.21	−0.06

relevant regions of space, the fractional Fukui function (and hence the variation of the density with N) is better modelled by RSH functionals with a value of $\mu \approx 0.4\, {a_0}^{-1}$ than by higher values of μ. This is true on both sides of integer N, despite a higher value of μ being more appropriate for the energy on the electron-deficient side.

Comment on electronegativity

Another quantity important to conceptual DFT is the electronegativity, which is related to the Fukui function, through the chemical potential, by equation (4.18). Mulliken's definition of the electronegativity, given in equation (4.14), can equivalently be approximated (*via* Janak's Theorem) as

$$\chi = -\frac{\varepsilon_{\mathrm{H}} + \varepsilon_{\mathrm{L}}}{2}\,, \tag{4.30}$$

which as we know will only truly hold in the absence of delocalisation error (although error-cancellation is possible). Given that this approximation involves the frontier orbital energies, we might expect the quality of the approximation to follow trends similar to the results of Chapter 3 when μ is varied, and *not* those of the Fukui function above. Table 4.1 compares errors in the electronegativity values from equation (4.14) with those from equation (4.30), for the same functional forms considered in Figure 4.4.

We established in Section 3.4.3 that I and A are, in general, in considerable error for the tuned RSH functionals—indeed this error contributes towards their success in estimating $\varepsilon_{\mathrm{L}} - \varepsilon_{\mathrm{H}}$. Thus when calculating χ using equation (4.14),

we would expect the tuned RSH functionals to give a poorer estimate of the electronegativity than conventional functionals. This is indeed the case, although increasing μ to $0.66\,{a_0}^{-1}$ reduces the error in A (the error in I is largely independent of μ) sufficiently to give an error in χ similar to that of CAM-B3LYP.

On the other hand, the tuned functionals show a considerable improvement over the non-tuned alternatives, when calculating χ using equation (4.30), due to the reduction in delocalisation error. The trends do not follow those observed for the fundamental gap in Chapter 3, since we are now dealing with the sum, rather than the difference, of the orbital energies, however there are some important observations. Specifically, the magnitude of the error decreases as μ—and hence the proportion of short-range exact exchange—increases. This contrasts with the trends exhibited by the Fukui function, and suggests that the optimal proportion for estimating the electronegativity from the orbital energies is much higher than for estimating its derivative with respect to the external potential.

4.4 Conclusions and further considerations

So far we have considered only a single system: carbon. However, even with such a limited test case, we have demonstrated that near-linearity in the energy variation does not necessarily correspond to near-linearity in the density variation with number of electrons. Near-linearity can be achieved in either property (for a given segment spanning two adjacent integer-N species), but each requires a different range-separation parameter.

The optimal value for density variation appears to correspond to a value typically chosen from an average over many thermochemical quantities and systems. It would be premature to draw any conclusions from only a single system, but further study could involve other systems to determine whether this observation remains true.

In principle, an optimal μ could be found for each property (such as the Fukui function, electronegativity, and others *ad infinitum*) and each system, in the spirit of Chapter 3, but with the knowledge that such values may

vary significantly, the idea quickly becomes overwhelming. In addition, the inability—in general—to achieve near-linearity for two pairs of integer-N species simultaneously makes application of the functional limited: being both system- and property-dependent it could only ever be utilised on a small scale, and this ultimately highlights limitations in the underlying functional form.

Instead, for the remainder of this thesis we focus our efforts towards addressing functional development at a more fundamental level, trialling simple forms that satisfy known exact behaviour. We begin by introducing some known mathematical relationships, build the functional forms, and then perform a variety of assessments to determine their success.

5

Development of novel explicit density functionals

In this chapter we present a novel approach to functional development, based on the homogeneity properties of explicit density functionals under density scaling. We begin by investigating the properties of a number of simple functional forms, to justify the use of density scaling as a tool for functional development. We then constrain the mathematical forms to approximately satisfy known exact scaling conditions, culminating in a self-consistent functional that approximately recovers the exact Koopmans condition $\varepsilon_{\text{H}} = -I$ associated with energy linearity. It also produces a potential with the correct asymptotic behaviour. The chapter concludes with a series of tests to highlight the successes and remaining challenges for this developmental process.

As established in Chapter 3, it is desirable to try to satisfy the exact Koopmans conditions

$$\varepsilon_{\text{H}} = -I_0 \tag{5.1}$$

and

$$\varepsilon_{\text{L}} = -A_0 \,, \tag{5.2}$$

which correspond to the slopes of the exact piecewise linear E *vs* N curve, on the electron-deficient and electron-abundant sides of the integer N_0, respec-

tively. These conditions—in which the terminology electron-deficient/abundant refers to the limiting behaviour as the electron number approaches N_0 from below/above—are appropriate for explicit density functionals and for orbital-dependent functionals within the usual generalised Kohn–Sham (GKS) formalism.[139]

With orbital-dependent functionals, namely global and range-separated hybrids, we have seen that it is possible to tune the parameters in order to approximately satisfy one or both of these conditions, although the tuning procedure still has a number of drawbacks, outlined in the final section of Chapter 3. For explicit density functionals, such as local functionals or generalised gradient approximations (GGAs), the situation is even less straightforward. Firstly, the lack of exact orbital exchange means there is no obvious parameter that can be systematically varied to adjust the HOMO and LUMO energies. Furthermore, for global hybrid and RSH functionals the non-multiplicative nature of the XC operator means equations (5.1) and (5.2) can both be approximately satisfied at the same time. For explicit density functionals, however, the XC potential is now multiplicative, and it is not in general possible to satisfy both conditions simultaneously.

This can be traced to the fact that the piecewise linearity of the exact energy leads to a jump in the exact XC potential by an amount Δ_{XC} as N increases through an integer. This jump—typically of several electron volts—is the integer discontinuity discussed in earlier chapters. The exact potential on the electron-deficient side of the integer does yield a HOMO energy that satisfies equation (5.1); however it is the *shifted* potential on the electron-abundant side of the integer that yields a LUMO energy that satisfies equation (5.2). Local functionals and GGAs are continuum approximations, and exhibit a potential that is continuous across the integer. In other words there is only one potential for the integer system, which cannot satisfy both conditions. In fact, these functionals yield a potential that approximately averages over the discontinuity, meaning HOMO energies are well above $-I_0$ and LUMO energies are well below $-A_0$.[149]

In this chapter, we present a novel approach to XC functional development based on density scaling considerations, culminating in an explicit density functional that obeys equation (5.1), and exhibits the correct asymptotic

behaviour in its potential. Our approach is based on approximately satisfying known exact scaling relations, so we begin by outlining the key aspects of the theory, and the influence of the integer discontinuity. Following this, we investigate the mathematical properties of simple functional forms in order to justify the consideration of density scaling in the development of functionals. We then impose constraints appropriate specifically for the limiting *electron-deficient* side of any integer-N system, to define two functional forms appropriate for this side of the integer. Crucially, these functionals obey the Koopmans condition of equation (5.1), and one additionally produces a potential with the correct asymptotic behaviour. The chapter concludes with an assessment of the performance of these functionals for the self-consistent calculation of a number of properties.

5.1 Scaling relations

In the context of DFT, scaling relations refer to the behaviour of a functional when some quantity is *scaled* in a particular manner. Most common is the concept of *coordinate scaling*,[242] where the coordinates are uniformly shrunk or dilated by a constant λ whilst preserving normalisation, such that the coordinate-scaled density is

$$\rho_\lambda(\boldsymbol{r}) = \lambda^3 \rho(\lambda \boldsymbol{r})\,. \tag{5.3}$$

If a functional of this scaled density is equivalent to the same functional of the *unscaled* density, multiplied by a constant λ^m,

$$F[\rho_\lambda] = \lambda^m F[\rho]\,, \tag{5.4}$$

then we say that the functional is homogeneous of degree m under coordinate scaling.

A similar condition can be identified if the density itself is scaled by a factor ξ, such that a functional is homogeneous of degree k under *density scaling* if it satisfies

$$F[\xi\rho] = \xi^k F[\rho]\,. \tag{5.5}$$

Equivalently, for $k \neq 0$,[243]

$$k = \frac{\int v_F(\boldsymbol{r})\rho(\boldsymbol{r})\,d\boldsymbol{r}}{F[\rho]}\,, \tag{5.6}$$

where $v_F(\boldsymbol{r}) = \delta F[\rho]/\delta\rho(\boldsymbol{r})$. Evaluation of k using equation (5.6) thus provides a mechanism for quantifying the behaviour of any functional $F[\rho]$ under density scaling. If the value of k is system-independent, then equation (5.5) is universally satisfied and the functional is homogeneous of degree k. If the value of k is system-dependent, then the functional is inhomogeneous; the degree of system-dependence then provides a measure of the degree of inhomogeneity.

In a recent study, Borgoo *et al.*[244] investigated the density scaling properties of the exact XC functional. The functional is, in fact, inhomogeneous, though following Zhao *et al.*[245] a system-dependent *effective homogeneity* can be defined, through equation (5.6) as

$$k_{\mathrm{XC}} = \frac{\int v_{\mathrm{XC}}(\boldsymbol{r})\rho(\boldsymbol{r})\,d\boldsymbol{r}}{E_{\mathrm{XC}}[\rho]}\,, \tag{5.7}$$

where $E_{\mathrm{XC}}[\rho]$ is the XC energy functional and $v_{\mathrm{XC}}(\boldsymbol{r}) = \delta E_{\mathrm{XC}}[\rho]/\delta\rho(\boldsymbol{r})$ is the XC potential. This quantity was evaluated in Ref. 244 for atoms and molecules at equilibrium geometries using near-exact XC potentials, electron densities and XC energies, determined from experimental and correlated *ab initio* data.

5.1.1 Influence of the integer discontinuity

A key consideration in the study of Borgoo *et al.* was the integer discontinuity in the exact XC potential. On the electron-deficient side of the integer, the exact potential, which we denote $v^{-}_{\mathrm{XC}}(\boldsymbol{r})$, decays to zero as $-1/r$ and yields the HOMO energy in equation (5.1). On the electron-abundant side, the exact potential is shifted at all points in space by the integer discontinuity Δ_{XC}, and yields the LUMO energy in equation (5.2); we denote this potential

$$v^{+}_{\mathrm{XC}}(\boldsymbol{r}) = v^{-}_{\mathrm{XC}}(\boldsymbol{r}) + \Delta_{\mathrm{XC}}\,. \tag{5.8}$$

The average of the two, which we denote $v^{\mathrm{av}}_{\mathrm{XC}}(\boldsymbol{r})$, is approximately equivalent to that exhibited by conventional local functionals and GGAs (except at long range, where these explicit density functionals decay exponentially to zero).

With three distinct potentials, equation (5.7) can be evaluated to produce

three near-exact effective homogeneities,

$$k_{\mathrm{XC}}^{-} = \frac{\int v_{\mathrm{XC}}^{-}(\boldsymbol{r})\rho(\boldsymbol{r})\,d\boldsymbol{r}}{E_{\mathrm{XC}}[\rho]}\,, \tag{5.9}$$

$$k_{\mathrm{XC}}^{+} = \frac{\int v_{\mathrm{XC}}^{+}(\boldsymbol{r})\rho(\boldsymbol{r})\,d\boldsymbol{r}}{E_{\mathrm{XC}}[\rho]} = k_{\mathrm{XC}}^{-} + \frac{N\Delta_{\mathrm{XC}}}{E_{\mathrm{XC}}[\rho]}\,, \tag{5.10}$$

and

$$k_{\mathrm{XC}}^{\mathrm{av}} = \frac{\int v_{\mathrm{XC}}^{\mathrm{av}}(\boldsymbol{r})\rho(\boldsymbol{r})\,d\boldsymbol{r}}{E_{\mathrm{XC}}[\rho]} = k_{\mathrm{XC}}^{-} + \frac{N\Delta_{\mathrm{XC}}}{2E_{\mathrm{XC}}[\rho]}\,. \tag{5.11}$$

The near-exact values for each of these quantities were investigated in Ref. 244: both k_{XC}^{-} and k_{XC}^{+} are relatively system-dependent, whereas $k_{\mathrm{XC}}^{\mathrm{av}}$ is close to 4/3 (the homogeneity of the Dirac exchange functional).

5.1.2 GENERAL FORM FOR A HOMOGENEOUS FUNCTIONAL

By generalising the work of Liu and Parr,[246] we introduce a generic explicit density functional form, which is homogeneous of degree m under coordinate scaling and homogeneous of degree k under density scaling for all n,

$$E[\rho] = \left[\int \rho(\boldsymbol{r})^{\frac{9k-nm}{9k-3m}}\, x^{n}\, d\boldsymbol{r}\right]^{\frac{3k-m}{3-n}}\,, \tag{5.12}$$

where x is the dimensionless reduced density gradient

$$x = \frac{|\nabla\rho|}{\rho^{\frac{4}{3}}}\,, \tag{5.13}$$

and n is a power governing the degree of gradient correction ($n = 0$ signifies a local functional). The functional can be multiplied by an arbitrary parameter without altering its homogeneity properties. For brevity, we introduce the notation "DS: k" to denote homogeneity of degree k under density scaling, and "CS: m" to denote homogeneity of degree m under coordinate scaling.

Although the form looks complicated, it simplifies readily for many common purposes—for example choosing $n = 0$ and $k = (3+m)/3$ removes the gradient dependence and reduces the external power to unity. Then, the functional with $m = 1$ yields $k = 4/3$, and recovers the Dirac exchange functional (when multiplied by the appropriate prefactor). Similarly, $m = 2$ recovers $k = 5/3$ and the Thomas–Fermi kinetic energy functional. We will return to

this general functional form a number of times throughout this chapter, as a tool for investigating the behaviour of functionals that exhibit specific scaling properties.

5.2 Combining Coulomb and XC terms

For a system with $0 < N \leqslant 1$ electrons, the Coulomb and exchange–correlation energies must cancel exactly in order to eliminate the one-electron SIE of equation (2.79). Zhang and Yang[143] determined that, given that $J[\rho]$ is DS: 2, $E_{\mathrm{XC}}[\rho]$ must also be DS: 2 on the electron-deficient side of the one-electron system. Otherwise, examination of equation (5.5) shows that even if $J[\rho] + E_{\mathrm{XC}}[\rho] \approx 0$ for *one* electron, the (approximate) equality will *not* hold for half, or any fraction, of an electron. In other words, the functional will still suffer from the delocalisation error (as is the case for conventional explicit density functionals, for example, which are approximately DS: 4/3).

In fact, Ref. 244 demonstrates that this is part of a larger problem: unlike $J[\rho]$, the exact $E_{\mathrm{XC}}[\rho]$ is *not* homogeneous, and its effective (electron-deficient) homogeneity k_{XC}^{-} is highly system-dependent. The *average* effective homogeneity $k_{\mathrm{XC}}^{\mathrm{av}}$, associated with a potential averaged over the integer discontinuity, is less system-dependent, and is in general close to 4/3 (note that conventional explicit density functionals—which approximately average over the discontinuity—are homogeneous of around this degree).

A crucial observation, though, is that the exact $J[\rho]$ and $E_{\mathrm{XC}}[\rho]$ behave differently under density scaling. It is interesting, then, to consider their *combined* homogeneity properties, and whether the two may be modelled as a single entity. In other words, we define an overall functional combining the Coulomb and XC functionals, which we term Ω,

$$\Omega[\rho] = J[\rho] + E_{\mathrm{XC}}[\rho]\,, \tag{5.14}$$

whose effective DS homogeneity is given by

$$k_\Omega = \frac{k_J J[\rho] + k_{\mathrm{XC}} E_{\mathrm{XC}}[\rho]}{\Omega[\rho]} \tag{5.15}$$

$$= \frac{2J[\rho] + k_{\mathrm{XC}} E_{\mathrm{XC}}[\rho]}{J[\rho] + E_{\mathrm{XC}}[\rho]}\,. \tag{5.16}$$

We desire a functional form that satisfies this homogeneity condition. If we consider the simple approximation

$$\Omega[\rho] = A[\rho](N-1)\,, \tag{5.17}$$

for any $A[\rho]$ that is DS: 1, we find that

$$k_\Omega = \frac{\int v_\Omega(\boldsymbol{r})\rho(\boldsymbol{r})\,d\boldsymbol{r}}{\Omega[\rho]} \tag{5.18}$$

$$= \frac{2N-1}{N-1}\,, \tag{5.19}$$

where $v_\Omega = \delta\Omega[\rho]/\delta\rho(\boldsymbol{r})$, and we have used the definition

$$k_A = \frac{\int v_A(\boldsymbol{r})\rho(\boldsymbol{r})\,d\boldsymbol{r}}{A[\rho]} = 1\,. \tag{5.20}$$

Importantly, the factor of $(N-1)$ ensures that $E_{\text{XC}}[\rho] = J[\rho]$ for a one-electron system.

Figure 5.1 plots the simple approximation of equation (5.19) against the near-exact k_Ω, computed using equation (5.16) with near-exact Zhao–Morrison–Parr (ZMP)[247] values of $E_{\text{XC}}[\rho]$ and $J[\rho]$, and the $k_{\text{XC}}^{\text{av}}$ values from Ref. 244. Note that we have regarded N as a functional of ρ in the differentiation; we could alternatively have chosen to treat N as a fixed parameter, which would result in a potential shifted only by a trivial additive constant (although the homogeneity of the functional would differ).

Even with such a simple approximation, the similarity in the trend is striking. We can add further physical justification to this form by noting that $J[\rho]/N$ is DS: 1 (when N is differentiated, as above). If we choose $A[\rho] = J[\rho]/N$, the overall functional becomes

$$\Omega[\rho] = \frac{J[\rho]}{N}(N-1) \tag{5.21}$$

$$= J[\rho]\left(1 - \frac{1}{N}\right)\,, \tag{5.22}$$

which is simply the Fermi–Amaldi approximation.[248] Whilst, in general, this is a poor model for the exchange–correlation functional, it has a number of desirable properties[249]—in particular it is exact for one-electron systems (*i.e.* it is one-electron SIE-free), and its functional derivative $v_{\text{XC}}(\boldsymbol{r})$ gives the correct asymptotic behaviour.

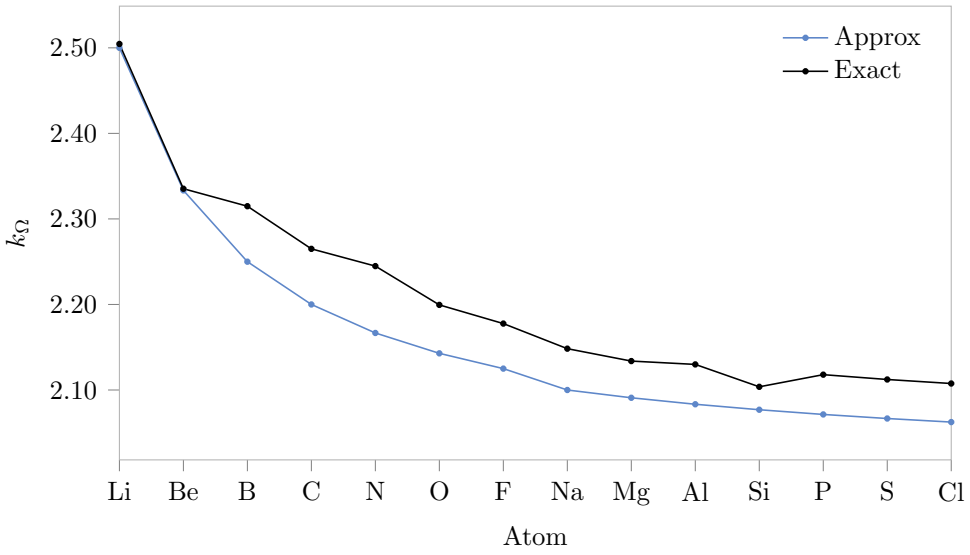

Figure 5.1: Comparison of exact k_Ω values from equation (5.16) with approximate values from equation (5.19).

It is interesting, then, to investigate whether we can improve upon this form with a simple correction term, based on the above scaling arguments. Let

$$A[\rho] = \frac{J[\rho]}{N} + B[\rho]\,, \tag{5.23}$$

where $B[\rho]$ (DS: 1) is our correction to Fermi–Amaldi. Thus

$$\Omega[\rho] = \left(\frac{J[\rho]}{N} + B[\rho]\right)(N-1)\,, \tag{5.24}$$

which—if we separate $\Omega[\rho]$ back into its constituent terms—corresponds to an exchange–correlation energy of

$$E_{\mathrm{xc}}[\rho] = -\frac{J[\rho]}{N} + B[\rho](N-1)\,. \tag{5.25}$$

For our choice of $B[\rho]$, we return to the generic form of equation (5.12), multiplied by a prefactor α, with our sole constraint being $k = 1$ (we do not know the exact coordinate-scaling behaviour). For simplicity, we initially choose $n = m$, such that the external power reduces to unity and we are left with

$$B[\rho] = \alpha \int \rho^{\frac{9-m^2}{9-3m}}\, x^m \, d\boldsymbol{r}\,. \tag{5.26}$$

A couple of important cases arise. Choosing $m = 0$ (hence CS: 0) reduces the functional simply to N multiplied by a constant. Choosing $m = 1$ (hence

Table 5.1: Calculated prefactors α computed using $B[\rho]$ in equation (5.26), along with mean absolute errors (MAE) and sums of square errors (SSE) in the resulting $E_{\mathrm{XC}}[\rho]$ values, compared to near-exact ZMP values. Fermi–Amaldi (FA) errors are included for comparison.

Functional	α	MAE/E_{h}	SSE/E_{h}
FA	—	6.95	1044.72
$m = 0$	-6.17×10^{-2}	0.29	1.44
$m = 1/3$	-3.53×10^{-2}	0.42	2.99
$m = 2/3$	-1.83×10^{-2}	0.59	5.99
$m = 3/3$	-8.56×10^{-3}	0.80	11.04
$m = 4/3$	-3.66×10^{-3}	1.04	18.50
$m = 5/3$	-1.45×10^{-3}	1.28	28.44
$m = 6/3$	-5.39×10^{-4}	1.54	40.68
$m = 7/3$	-1.92×10^{-4}	1.78	54.90
$m = 8/3$	-6.61×10^{-5}	2.02	70.67
$m = 9/3$	-2.23×10^{-5}	2.24	87.60
$m = 10/3$	-7.38×10^{-6}	2.45	105.30
$m = 11/3$	-2.42×10^{-6}	2.64	123.47
$m = 12/3$	-7.86×10^{-7}	2.82	141.84

CS: 1) reduces the exponent to 4/3, thus mimicking Dirac exchange with a gradient correction.

5.2.1 TESTING THE FUNCTIONAL

We implemented the functional in a post-Kohn–Sham manner into a development version of the CADPAC program, using the converged PBE density, for a range of m between 0 and 4. The parameter α was then determined by a least-squares fit of the RHS of equation (5.25) (using the near-exact ZMP[247] values of $J[\rho]$) to the near-exact ZMP values of $E_{\mathrm{XC}}[\rho]$, for the atoms Li–F and Na–Cl. Calculations were performed using the aug-cc-pVTZ basis set.

Table 5.1 lists the prefactors α, determined from the least-squares fit, for the range of $B[\rho]$ functionals tested, along with the mean absolute errors (MAE) and sum of square errors (SSE) in the resulting $E_{\mathrm{XC}}[\rho]$, compared to near-exact ZMP values. Errors in the unmodified Fermi–Amaldi form are included for comparison.

It is immediately obvious that all of the functional forms tested provide a significant improvement over Fermi–Amaldi, with the mean absolute error over the set of atoms reducing from almost $7\,E_{\mathrm{h}}$ to a little as $0.3\,E_{\mathrm{h}}$. It is not really

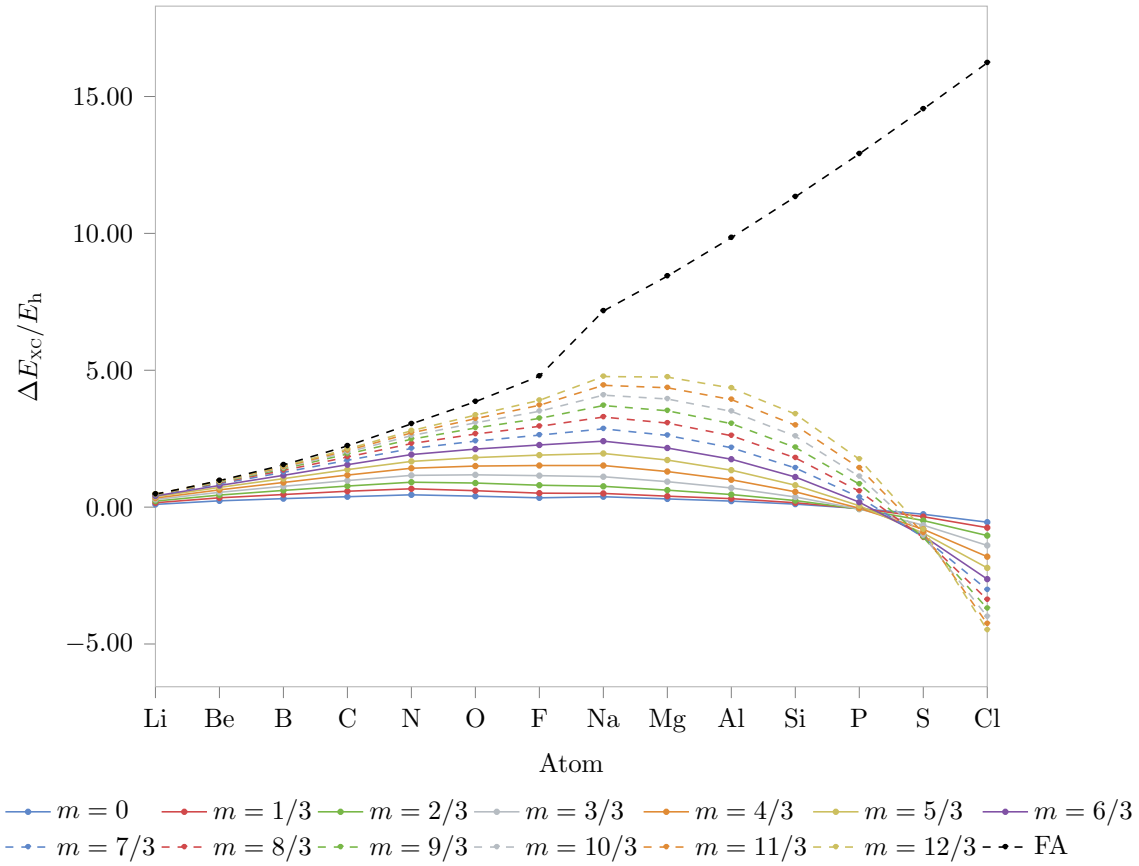

Figure 5.2: Errors in $E_{\mathrm{XC}}[\rho]$, compared to near-exact ZMP values, for a range of $B[\rho]$ functionals, where $m = n$, along with unmodified Fermi–Amaldi (FA).

surprising that the errors in Fermi–Amaldi are so high, since it was devised for the useful properties of its functional derivative rather than its quality as an approximation for E_{XC}. What is perhaps more surprising is that adding such a simple correction has such a pronounced effect—indeed it is the simplest correction, $E_{\mathrm{XC}}[\rho] = -J[\rho]/N + \alpha N(N-1)$, which results in the lowest errors.

This is particularly noticeable in Figure 5.2, which plots the error in E_{XC} for each atom: whilst Fermi–Amaldi gets rapidly worse as the atom size increases, the corrected forms, in particular $m = 0$, remain small in error. It must be noted that the increased error in the corrected forms with higher m is likely to do with the magnitude of $B[\rho]$ prior to scaling by α. It can be seen from the increasingly small values of α required that as m increases, so does the

magnitude of $B[\rho]$, and so one might expect a larger spread of individual errors.

5.2.2 BREAKING THE DENSITY-SCALING REQUIREMENT

So far, we have only considered functional forms which are homogeneous of degree one under density scaling, due to the similarity this produces in the overall effective homogeneity to the near-exact values in Figure 5.1. We carried out a brief investigation into breaking this DS: 1 "requirement" as a possible means of further improving the $B[\rho] = \alpha N$ correction.

We considered a functional $B[\rho] = \alpha N^p$, to give

$$\Omega[\rho] = \left(\frac{J[\rho]}{N} + \alpha N^p \right) (N - 1) \,, \tag{5.27}$$

such that the effective homogeneity is

$$k_\Omega = \frac{\int v_\Omega \rho \, d\boldsymbol{r}}{\Omega[\rho]} = \frac{(J[\rho]/N + \alpha p N^p)(N-1) + J[\rho] + \alpha N^{p+1}}{(J[\rho]/N + \alpha N^p)(N-1)} \,, \tag{5.28}$$

and substituting $p = 1$ recovers

$$k_\Omega = \frac{2N - 1}{N - 1} \,. \tag{5.29}$$

For the purposes of our investigation we may compute this functional—and the corresponding E_{XC}—entirely from near-exact ZMP quantities and a knowledge of the number of electrons in the system. The prefactor α and exponent p are again determined from a fit over the set of atoms. There are, in fact, two intuitive quantities we can fit to. Based on our above homogeneity arguments we could choose the parameters which minimise the difference between k_Ω in equation (5.28), and the near exact value in equation (5.16). Alternatively, we could choose the parameters that minimise the error in E_{XC}. In the ideal case the two approaches would give the same value, but in the event that this is *not* the case, we could attempt to fit to both quantities simultaneously. We take two approaches to the fitting: either minimising the sum of square errors (SSE) as before, or minimising the mean absolute percentage error (MAPE); the latter method was chosen to better weight the two components when fitting to both k_Ω and E_{XC} simultaneously.

Table 5.2 summarises the results of the different minimisation criteria. It is clearly possible to improve upon the description of both E_{XC} and k_Ω by

Table 5.2: Optimised parameters for the functional $B[\rho] = \alpha N^p$, along with the errors in k_Ω and E_{XC} (compared to near-exact ZMP values), generated for various minimisation conditions. Both mean absolute percentage errors (MAPE) and sums of square errors (SSE) are quoted in E_h.

			k_Ω		E_{XC}	
Minimisation criteria	α	p	MAPE	SSE	MAPE	SSE
SSE						
E_{XC} with $p = 1$	−0.062	1	0.020	0.034	0.038	1.442
E_{XC}	−0.104	0.805	0.063	0.284	0.004	0.019
k_Ω	−0.097	0.958	0.006	0.006	0.151	167.735
E_{XC} and k_Ω	−0.094	0.843	0.042	0.127	0.008	0.077
MAPE						
E_{XC} with $p = 1$	−0.064	1	0.020	0.034	0.034	2.826
E_{XC}	−0.108	0.792	0.071	0.357	0.004	0.026
k_Ω	−0.093	0.951	0.006	0.006	0.119	110.610
E_{XC} and k_Ω	−0.077	0.927	0.007	0.007	0.021	1.072

breaking the DS: 1 requirement of our (very simple) functional form, although reassuringly the optimal values remain relatively close to unity. There is a certain ambiguity as to which value is "best", but it suffices to note that the errors in E_{XC} and k_Ω are minimised for different values of p: approximately 0.8 and 0.95, respectively. A good compromise between the two can be found by minimising the MAPE of both simultaneously, with $p \approx 0.93$.

Figures 5.3 and 5.4 plot k_Ω and the error in E_{XC} respectively, for the atoms involved in the fit. The trends clearly reinforce those of Table 5.2—individually optimising k_Ω or E_{XC} negatively affects the other quantity, but optimising both simultaneously provides a reasonable (if not perfect) approximation of both. Constraining p to be unity, as in our initial tests, also gives a reasonable approximation.

This analysis is very promising. By starting from basic homogeneity arguments, we have taken an exceedingly simple functional form—N raised to a power and multiplied by a constant—and added it to the Fermi–Amaldi functional to produce a remarkable correction, whereby the correct exchange–correlation energies and effective homogeneities are approximately reproduced. The functional form is not perfect—and, indeed, we could continue to look for a more optimal set of parameters in equation (5.26) by removing our simplistic constraints, however there are other considerations to take into account.

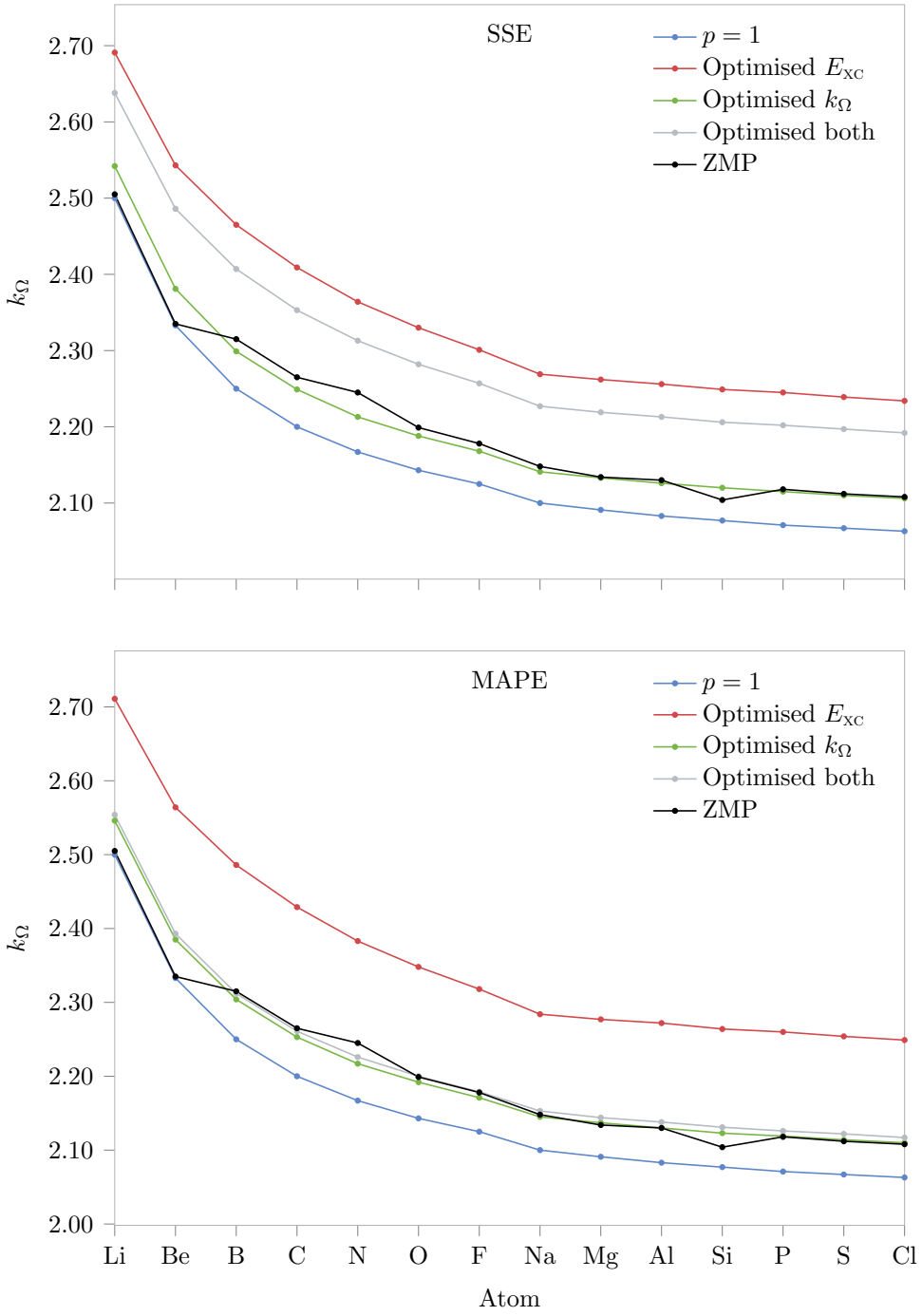

Figure 5.3: k_Ω values for $B[\rho] = \alpha N^p$ using different optimisation criteria, compared with near-exact values. SSE indicates the optimisations were carried out by minimising the sum of square errors, whilst MAPE indicates minimisation of the mean absolute percentage error.

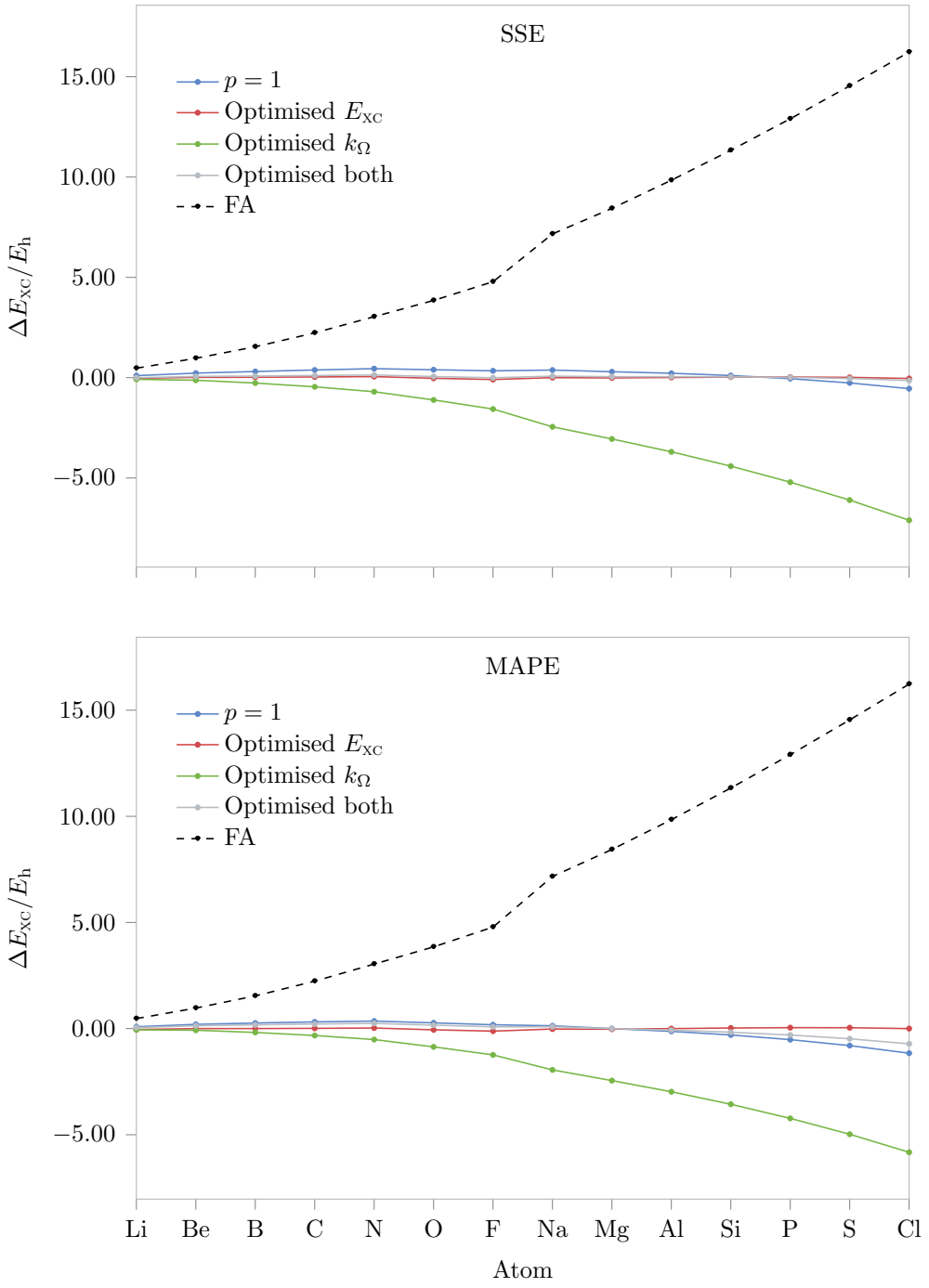

Figure 5.4: Errors in $E_{\mathrm{XC}}[\rho]$ with respect to near-exact ZMP values, for $B[\rho] = \alpha N^p$ using different optimisation criteria, along with unmodified Fermi–Amaldi (FA). SSE indicates the optimisations were carried out by minimising the sum of square errors, whilst MAPE indicates minimisation of the mean absolute percentage error.

Our investigations so far have originated from the qualitative observation in Figure 5.1 that a simple mathematical form mimics k_Ω, as defined by equation (5.16) when evaluated using $k_{\text{XC}}^{\text{av}}$, from equation (5.11). We have demonstrated that we can use this to derive a functional form that, through exhibiting the correct effective homogeneity, can provide a good approximation to the exchange–correlation energy.

We note, however, that the underlying homogeneity constraint ($k_{\text{XC}} = k_{\text{XC}}^{\text{av}}$) relates to the exchange–correlation potential *averaged* across the integer discontinuity. Thus, constraining a functional to this homogeneity will by definition generate a potential of this averaged, GGA-like nature, and as such *will not solve our overarching concern—that of the delocalisation error and satisfying the Koopmans conditions of equations* (5.1) *and* (5.2).

As discussed in the introduction to this chapter, the continuum nature of explicit density functionals means we cannot simultaneously satisfy both equations (5.1) and (5.2). Instead we now consider using the above homogeneity arguments to develop an expression appropriate specifically for the electron-deficient side of the integer system, culminating in a functional that *does* approximately satisfy equation (5.1).

For the remainder of the chapter, then, we look to replacing $k_{\text{XC}}^{\text{av}}$ in our functional form with k_{XC}^{-}. As noted in Ref. 244, the system-dependence of k_{XC}^{-} is much more pronounced than that of $k_{\text{XC}}^{\text{av}}$, and finding an appropriate approximation may be non-trivial. However, by relating the properties of a functional studied by Parr and Ghosh[250] to the homogeneity calculations of Borgoo *et al.*,[244] we establish that the necessary system-dependence can be introduced directly into the generic functional of equation (5.12), rather than by an equivalent of the $(N-1)$ term in equation (5.25).

5.3 Parr–Ghosh functional

The functional investigated by Parr and Ghosh[250] takes a form very similar to equation (5.25), neglecting only the factor of $(N-1)$,

$$E_{\text{XC}}[\rho] = -\frac{J[\rho]}{N} + G[\rho]\,, \tag{5.30}$$

where $G[\rho]$ is an arbitrary functional which is homogeneous of degree one under density scaling. Whilst such a homogeneity condition is acknowledged in the above paper as a simple approximation, it provides an excellent basis for our own investigation of homogeneity relationships.

From equation (5.30), we can determine the exchange–correlation potential

$$v_{\mathrm{XC}}(\boldsymbol{r}) = \frac{\delta}{\delta\rho}\left(-\frac{J[\rho]}{N}\right) + \frac{\delta}{\delta\rho}\big(G[\rho]\big) \tag{5.31}$$

$$= -\frac{v_J[\rho]}{N} + v_G[\rho]\,, \tag{5.32}$$

and the effective homogeneity

$$k_{\mathrm{XC}}^{\mathrm{PG}} = \frac{\int v_{\mathrm{XC}}(\boldsymbol{r})\rho(\boldsymbol{r})\,d\boldsymbol{r}}{E_{\mathrm{XC}}[\rho]} \tag{5.33}$$

$$= \frac{-2J[\rho]/N + G[\rho]}{E_{\mathrm{XC}}[\rho]} \tag{5.34}$$

$$= \frac{E_{\mathrm{XC}}[\rho] - (J[\rho]/N)}{E_{\mathrm{XC}}[\rho]}\,, \tag{5.35}$$

where we have used the fact that $G[\rho]$ is DS: 1,

$$k_G\,G[\rho] = G[\rho] = \int v_G(\boldsymbol{r})\,\rho(\boldsymbol{r})\,d\boldsymbol{r}\,. \tag{5.36}$$

Note that, following Parr and Ghosh, we have *not* differentiated N, treating it instead as a fixed parameter. This can be rationalised by considering the nature of the electron–deficient potential: by not differentiating N, the Parr–Ghosh potential, being a correction to Fermi–Amaldi, correctly decays to zero as $-1/r$. If we instead chose to differentiate N, the shape of the potential would be unchanged, but would be shifted by a constant and no longer decay to zero.

For the atoms for which we have near-exact ZMP quantities, we can hence determine the effective homogeneities of the Parr–Ghosh functional. Table 5.3 presents the deviations in these quantities from the near-exact k_{XC}^{-} values, and correlates them with the errors in the Parr–Ghosh functional. Immediately, we see that for atoms where the error in the energy is small, the deviation in $k_{\mathrm{XC}}^{\mathrm{PG}}$ is also small, indicating that constraining $G[\rho]$ to be DS: 1 is a good approximation. Conversely, however, the atoms for which the Parr–Ghosh approximation breaks down and the error in the energy is large, are those where $k_{\mathrm{XC}}^{\mathrm{PG}}$ shows a significant error, implying that, for these systems, k_G should not be unity.

Table 5.3: Deviation of the effective homogeneities of the Parr–Ghosh (PG) functional from near-exact k_{XC}^{-}, along with the errors in total energy (from near-exact values) quoted in Ref. 250.

Atom	$k_{\mathrm{XC}}^{\mathrm{PG}} - k_{\mathrm{XC}}^{-}$	$\Delta E/E_{\mathrm{h}}$
Li	0.217	0.398
Be	0.045	0.149
B	0.044	0.177
C	0.003	0.010
N	−0.024	−0.244
O	−0.001	−0.020
F	−0.008	−0.082
Na	0.113	1.609
Mg	0.084	1.585
Al	0.059	1.101
P	−0.008	−0.165
S	0.002	0.016
Cl	−0.011	−0.350

Can we, therefore, improve upon the Parr–Ghosh approximation by choosing an appropriate $G[\rho]$ with the correct homogeneity? In other words, if we choose a functional form for $G[\rho]$ by constraining k_G such that $k_{\mathrm{XC}} = k_{\mathrm{XC}}^{-}$ can we reduce the errors in Table 5.3? Moreover, since the homogeneity is intrinsically linked to the potential through equation (5.9), can we produce a functional that satisfies $\varepsilon_{\mathrm{H}} = -I$?

5.4 Development of an electron-deficient functional

We present two functional forms, where the effective DS homogeneity is constrained to be k_{XC}^{-}; we will term these functionals ED, for electron-deficient. The first approach, termed ED1, is given by

$$E_{\mathrm{XC}}^{\mathrm{ED1}}[\rho] = \gamma F[\rho]\,, \tag{5.37}$$

which is a simple scaling of an explicit density functional. The second approach, termed ED2, follows the Parr–Ghosh method to provide a correction to the Fermi–Amaldi approximation, whereby

$$E_{\mathrm{XC}}^{\mathrm{ED2}}[\rho] = -\frac{J[\rho]}{N} + \alpha G[\rho]\,. \tag{5.38}$$

$F[\rho]$ and $G[\rho]$ are functionals of the form equation (5.12), which are DS: k and CS: m for all n, and γ and α are arbitrary constants. The latter approach, as we have seen, is desirable due to the correct asymptotic behaviour of the potential—though for completeness we will investigate both methods.

In both cases, we desire an exchange–correlation functional with an effective homogeneity of $k_{\text{XC}} = k^-_{\text{XC}}$, and so for ED1 this requires

$$k_F = k^-_{\text{XC}} \,. \tag{5.39}$$

For ED2, we evaluate equation (5.7) for the functional in equation (5.38), and substitute $k_{\text{XC}} = k^-_{\text{XC}}$, to give

$$k^-_{\text{XC}}\, E_{\text{XC}}[\rho] = \int v_{\text{XC}}(\boldsymbol{r})\rho(\boldsymbol{r})\, d\boldsymbol{r} \tag{5.40}$$

$$= \int \frac{-v_J(\boldsymbol{r})}{N}\rho(\boldsymbol{r})\, d\boldsymbol{r} + \alpha \int v_G(\boldsymbol{r})\, \rho(\boldsymbol{r})\, d\boldsymbol{r} \tag{5.41}$$

$$= \frac{-2J[\rho]}{N} + \alpha\, k_G\, G[\rho] \,. \tag{5.42}$$

This rearranges, dropping $[\rho]$ for brevity, to give

$$k_G = \frac{k^-_{\text{XC}} E_{\text{XC}} + (2J/N)}{E_{\text{XC}} + (J/N)} \,. \tag{5.43}$$

As before, we have treated N as a fixed parameter. We then substitute k_F or k_G from equations (5.39) and (5.43) for k in the generic functional equation (5.12), to give the functionals $F[\rho]$ and $G[\rho]$, respectively.

Table 5.4 lists the required effective homogeneities for a range of atoms and molecules for each approach. Quantities were derived from near-exact ZMP values where available, or otherwise from the near-exact calculations of Ref. 244. Consistent with our observations in Section 5.3, the k_G values that are close to unity in Table 5.4 correspond to atoms—for example oxygen, carbon and sulphur—where the Parr–Ghosh approximation was particularly effective (*i.e.* small errors in Table 5.3).

Given that k is pre-determined, we now have three free parameters for which we must find optimal values: the coordinate-scaling m, the prefactor γ or α, and the gradient exponent n. To begin with, for the sake of simplicity, we choose to eliminate the gradient dependence and generate a local functional where $n = 0$, so that equation (5.12) becomes

$$E[\rho] = \left[\int \rho(\boldsymbol{r})^{\frac{3k}{3k-m}}\, d\boldsymbol{r}\right]^{\frac{3k-m}{3}} \,, \tag{5.44}$$

Table 5.4: Exact effective homogeneities k_F and k_G required for functionals $F[\rho]$ and $G[\rho]$, in order to ensure $k_{\text{XC}} = k_{\text{XC}}^{-}$, for equations (5.37) and (5.38).

Atom	k_F	k_G
Li	1.525	0.161
Be	1.606	0.867
B	1.558	0.889
C	1.568	0.993
N	1.575	1.053
O	1.544	1.003
F	1.545	1.017
Ne	1.547	1.035
Na	1.391	0.772
Mg	1.403	0.835
Al	1.410	0.888
P	1.451	1.015
S	1.431	0.997
Cl	1.435	1.019
Ar	1.440	1.037

Molecule	k_F	k_G
CH_4	1.552	1.280
CO	1.486	1.151
Cl_2	1.410	1.211
F_2	1.501	1.233
H_2O	1.513	1.017
HCl	1.428	1.030
HF	1.523	1.011
N_2	1.494	1.165

replacing k with the exact k_F or k_G from Table 5.4 to produce $F[\rho]$ and $G[\rho]$.

5.4.1 POST-KOHN–SHAM OPTIMISATION

As before, we used a post-KS framework in a developmental version of CADPAC to determine $F[\rho]$ and $G[\rho]$, using the converged PBE density with the aug-cc-pVTZ basis set. Equation (5.12) was evaluated for $0.9 \leqslant m \leqslant 2$ at intervals of 0.05; n was set to zero and k taken from Table 5.4.

For each m, we then performed a least-squares fit, to near-exact values, of $F[\rho]$ against $E_{\text{XC}}[\rho]$, and $G[\rho]$ against $E_{\text{XC}}[\rho] + J[\rho]/N$, in order to determine γ and α, respectively. The value of m that produced the lowest errors in the fit was chosen, along with its corresponding prefactor, as the optimal functional form. The fitting set comprised the species in Table 5.4; we investigated fitting to both the set of atoms only, in keeping with previous work, and to the combined set of atoms and molecules.

Figure 5.5 plots the regression fit to determine the prefactor for the optimal m of each of the four approaches. In plot (a), fitting only to the set of atoms, there is a stark contrast between $F[\rho]$ and $G[\rho]$. The former shows a distinct jump in the correlation between the computed and near-exact values, consistent

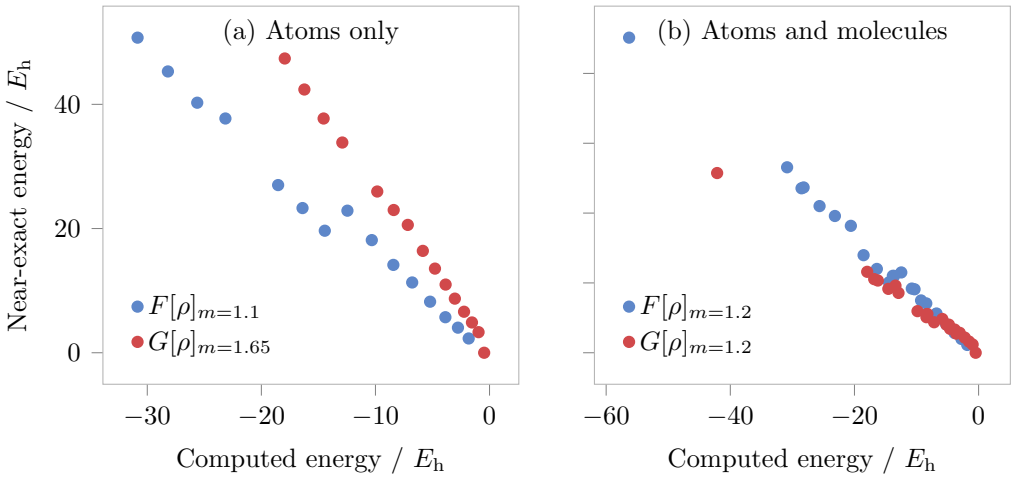

Figure 5.5: Linear-regression fits (through zero) of the ED approximate energy functionals against near-exact values, to determine the prefactors γ and α. Fits were performed (a) to the set of atoms in Table 5.4 and (b) to both the atoms *and* molecules in the same table.

with the periodicity of the atoms (the jump occurs between Ne and Na). The nature of this jump means a simple linear relationship cannot be found and the errors due to the averaged prefactor γ are significant. $G[\rho]$, however, appears to have a much more linear correspondence with the near-exact values, with the periodicity largely accounted for by the $J[\rho]/N$ term. As a result, the averaged prefactor α produces surprisingly small errors across the range of atoms, and is an exciting prospect for such a simple functional form.

Plot (b) includes the molecules in the fit, and the difference between the two functionals is now lessened by the increasingly scattered correlation. It is perhaps to be expected that the introduction of molecular species reduces the possibility of finding a simple linear relationship between our functional forms and the near-exact value, however the fact they are at all close is encouraging.

Table 5.5 lists the optimal parameters, as determined by the regression fits, for the four functionals. For notational purposes, we use the suffixes -A and -AM to denote the functionals fitted to atoms, and atoms and molecules, respectively.

Table 5.5: Parameters defining the "optimal" ED1 and ED2 functionals, as determined from a linear regression fit to the set of atoms (-A) and atoms and molecules (-AM).

Functional	γ	m
ED1-A	−0.629 15	1.10
ED1-AM	−0.610 97	1.20
	α	m
ED2-A	−0.376 89	1.65
ED2-AM	−0.785 30	1.20

5.4.2 SELF-CONSISTENT IMPLEMENTATION

With all parameters now defined, the ED1 and ED2 functionals were implemented self-consistently into the developmental CADPAC code. The exchange–correlation potentials were obtained directly by functionally differentiating all terms in equations (5.37) and (5.38), given the relevant constants. Note that the ED2 functional must take into account the derivative of the $-J/N$ term, for which we treat N as a fixed constant.

For convenience, we define $p = 3k/(3k-m)$, $q = (3k-m)/3$, and $\varepsilon[\rho] = \int \rho^p \, d\boldsymbol{r}$, and so the generic energy functional in equation (5.44) becomes

$$E[\rho] = \left[\int \rho^p \, d\boldsymbol{r} \right]^q = \varepsilon[\rho]^q \,. \tag{5.45}$$

The potential is then

$$\begin{aligned} v_E = \frac{\delta E[\rho]}{\delta \rho} &= q\varepsilon[\rho]^{q-1} \frac{\delta \varepsilon[\rho]}{\delta \rho} && (5.46)\\ &= pq\varepsilon[\rho]^{q-1} \, \rho^{p-1} && (5.47)\\ &= k\varepsilon[\rho]^{\left(k-\frac{m}{3}-1\right)} \, \rho^{\frac{m}{3k-m}} \,. && (5.48) \end{aligned}$$

Special care must be taken due to the external power q in equation (5.44), as the energy functional appears—raised to a different power—in the expression for the potential, and this must be properly accounted for in the implementation.

The TDDFT exchange–correlation kernel was similarly determined from the second functional derivative of equation (5.45),

$$\begin{aligned} f_{\mathrm{XC}}^{\mathrm{ED}} = \frac{\delta^2 E[\rho]}{\delta\rho(\boldsymbol{r})\delta\rho(\boldsymbol{r}')} &= qp(p-1)\varepsilon[\rho]^{q-1} \, \rho(\boldsymbol{r})^{p-2} \delta(\boldsymbol{r}-\boldsymbol{r}') \\ &\quad + p^2 q(q-1)\varepsilon[\rho]^{q-2} \, \rho(\boldsymbol{r})^{p-1} \, \rho(\boldsymbol{r}')^{p-1} \,. \end{aligned} \tag{5.49}$$

Extra care must again be taken in the evaluation of the two-electron integral $(ia|f_{\text{XC}}^{\text{ED}}|jb)$ from equation (2.78), as the second term of equation (5.49) depends on both $\boldsymbol{r}$ and $\boldsymbol{r}'$, which again arises due to the external power q in equation (5.45). Expansion of the integrals leads to two terms,

$$\begin{aligned}\iint d\boldsymbol{r}\, d\boldsymbol{r}' \, \frac{\delta^2 E[\rho]}{\delta\rho(\boldsymbol{r})\delta\rho(\boldsymbol{r}')}\, \varphi_i(\boldsymbol{r})\varphi_a(\boldsymbol{r})\varphi_j(\boldsymbol{r}')\varphi_b(\boldsymbol{r}') \\ = qp(p-1)\varepsilon[\rho]^{q-1} \int d\boldsymbol{r}\, \rho(\boldsymbol{r})^{p-2}\, \varphi_i(\boldsymbol{r})\varphi_a(\boldsymbol{r})\varphi_j(\boldsymbol{r})\varphi_b(\boldsymbol{r}) \\ + p^2 q(q-1)\varepsilon[\rho]^{q-2} \int d\boldsymbol{r}\, \rho(\boldsymbol{r})^{p-1}\, \varphi_i(\boldsymbol{r})\varphi_j(\boldsymbol{r}) \int d\boldsymbol{r}'\, \rho(\boldsymbol{r}')^{p-1} \varphi_j(\boldsymbol{r}')\varphi_b(\boldsymbol{r}')\,. \end{aligned} \tag{5.50}$$

A full derivation can be found in Appendix A.

Verifying the implementation

Our implementation was verified by carrying out Hellmann–Feynman tests, comparing analytic results for first and second derivatives (the dipole moment and static polarisability, respectively) with numerical finite difference results (using energies and dipole moments, respectively) determined from calculations in a small electric field. Table 5.6 gives a representative example of this process for the CO molecule, using the full system-dependent form of the ED2 functional (defined in Section 5.5) and the Sadlej basis set used in the analysis of static polarisabilities (Section 5.6.4). There is close agreement between the numerical and analytic results, comparable to equivalent tests performed using PBE (the results of which are included in Table 5.6 for comparison), verifying that the implementation is variational. Note that the parameters defining each of the functionals were treated as independent of electric field.

Performance of the self-consistent functionals

For simplicity, the self-consistent implementation was restricted to closed-shell systems. Although this limited the available set of test systems, it was sufficient to draw initial conclusions.

Table 5.7 lists the mean absolute errors (MAEs) in E_{XC} (compared to near-exact ZMP values) and those in the HOMO energies ε_{H} (compared to $-I_0$

Table 5.6: Representative Hellmann–Feynman tests for CO, using the ED2 functional and Sadlej basis set, comparing the agreement of analytic dipole moments μ and static isotropic polarisabilities $\alpha_{\text{iso}} = (\alpha_{xx}+\alpha_{yy}+\alpha_{zz})/3$, with numerical equivalents. Numerical quantities were determined by finite difference quotients from calculations in an electric field of ± 0.0001 a.u., and all quantities are quoted in atomic units. PBE results are included for comparison.

Field	Energy in +ve field	Energy in -ve field	Difference quotient
x	−113.237 500	−113.237 500	0
y	−113.237 500	−113.237 500	0
z	−113.237 493	−113.237 506	0.0651
		Numerical μ_z	6.5135×10^{-2}
		Analytic μ_z	6.5132×10^{-2}
		Agreement	2.7800×10^{-6}
		PBE agreement	2.0900×10^{-6}
Field	Dipole in +ve field	Dipole in -ve field	Difference quotient
x	0.001 075	−0.001 075	10.7451
y	0.001 075	−0.001 075	10.7451
z	−0.063 763	−0.066 502	13.6958
		Numerical α_{iso}	11.7287
		Analytic α_{iso}	11.7287
		Agreement	2.9360×10^{-5}
		PBE agreement	1.1607×10^{-4}

values) for the set of closed-shell molecules in Table 5.4—this time computed self-consistently.

Despite the functional parameters being fitted (post-KS) to near-exact ZMP E_{XC} values, the errors in the *self-consistent* E_{XC} values are not particularly impressive for any of the functionals—at best they are an order of magnitude worse than PBE.

Fitted only to the the set of atoms ED2-A produces an MAE in E_{XC} of over $3\,E_{\text{h}}$—three times larger that of ED1-A, despite a much more linear trend in the regression fit of ED2-A. This seemingly incongruous result can be rationalised by the fact that this much more accurate fit is to a set of largely open-shell atoms, resulting in a functional that is completely inappropriate for closed-shell molecules. In other words, the poor fit of ED1-A to the set of atoms actually improves its application to molecules—although results are still poor.

When molecules are included in the fitting set, the MAEs in E_{XC} reduce for both methods, although they remain significant: ED1-AM produces an MAE

Table 5.7: Mean absolute errors in the exchange–correlation and HOMO energies, for self-consistent calculations using ED1 and ED2, with optimal parameters determined from the linear-regression fits of Figure 5.5. PBE and LDA errors are included for comparison.

Functional	$(E_{\mathrm{XC}} - E_{\mathrm{XC}}^{\mathrm{ZMP}})/E_{\mathrm{h}}$	$(\varepsilon_{\mathrm{H}} + I_0)/E_{\mathrm{h}}$
ED1-A	0.987	0.113
ED2-A	3.180	0.141
ED1-AM	0.593	0.206
ED2-AM	1.160	0.067
PBE	0.087	0.189
LDA	1.202	0.194

of nearly $0.6\,E_{\mathrm{h}}$ (over six times larger than PBE), whilst ED2-AM gives an error double this again. These large errors are symptomatic of the significant scattering around the linear fit.

The errors in ε_{H} are more promising, with both ED1 and ED2 showing an improvement over PBE and LDA. In particular, ED2-AM gives a threefold reduction in the error compared to PBE, however, the large errors in E_{XC} for this functional limits its usefulness—there is little benefit to improving the HOMO energies if the overall energies are ruined.

Two problems thus remain with our functional form. Firstly, our theoretical justification for these functional forms relies on explicitly constraining the homogeneity of the functional to a system-dependent k_{XC}^{-}, a quantity for which we have used known near-exact values from Ref. 244. In the general case, however, this quantity is unknown, rendering the functional form useless without prior knowledge of exact data or highly accurate *ab-initio* calculations. Secondly, we have not succeeded in finding a prefactor that is appropriate for all systems—the energies simply do not give a sufficiently linear relationship with our functional form. We address the first problem in the following section, and then move on to show how the solution can also elegantly solve the latter, to create a truly general, system-dependent functional form with promising characteristics.

5.4.3 Approximating the effective homogeneity

To turn our functional forms into a practical, generally applicable method, we need to be able to evaluate k for an arbitrary system, which poses an

obvious challenge given the form of equations (5.9), (5.39) and (5.43). However, we observe that all of the relevant components in these equations can be estimated relatively accurately from conventional GGA calculations, meaning an approximate k can be calculated in advance and then used in the functional. We do this as follows.

The quantities E_{XC} and J are trivially approximated from a GGA calculation; we denote them $E_{\mathrm{XC}}^{\mathrm{GGA}}$ and J^{GGA}. For k_{XC}^{-}, we first rearrange equation (5.11) to give

$$k_{\mathrm{XC}}^{-} = k_{\mathrm{XC}}^{\mathrm{av}} - \frac{N\Delta_{\mathrm{XC}}}{2E_{\mathrm{XC}}}\,, \tag{5.51}$$

noting from Ref. 244 that $k_{\mathrm{XC}}^{\mathrm{av}} \approx 4/3$. We also note that Δ_{XC} can be approximated by

$$\Delta_{\mathrm{XC}} \approx 2(\varepsilon_{\mathrm{H}}^{\mathrm{GGA}} + I^{\mathrm{GGA}})\,, \tag{5.52}$$

where $\varepsilon_{\mathrm{H}}^{\mathrm{GGA}}$ and I^{GGA} are the HOMO energy and ΔSCF ionisation potential (computed from total electronic energies) determined using a GGA functional. Equation (5.52) is central to the asymptotic approach of Ref. 191; see Refs 149 and 159 for further discussion. Substituting these two results into equation (5.51), and estimating E_{XC} using the GGA value, gives

$$k_{\mathrm{XC}}^{-} \approx \frac{4}{3} - \frac{N(\varepsilon_{\mathrm{H}}^{\mathrm{GGA}} + I^{\mathrm{GGA}})}{E_{\mathrm{XC}}^{\mathrm{GGA}}}\,. \tag{5.53}$$

Returning to equations (5.39) and (5.43), we therefore obtain the following expressions for the approximate, system-dependent k:

$$k_F \approx \frac{4}{3} - \frac{N(\varepsilon_{\mathrm{H}}^{\mathrm{GGA}} + I^{\mathrm{GGA}})}{E_{\mathrm{XC}}^{\mathrm{GGA}}}\,, \tag{5.54}$$

and

$$k_G \approx \frac{\left(\frac{4}{3} - \frac{N(\varepsilon_{\mathrm{H}}^{\mathrm{GGA}} + I^{\mathrm{GGA}})}{E_{\mathrm{XC}}^{\mathrm{GGA}}}\right) E_{\mathrm{XC}}^{\mathrm{GGA}} + (2J^{\mathrm{GGA}}/N)}{E_{\mathrm{XC}}^{\mathrm{GGA}} + (J^{\mathrm{GGA}}/N)}\,. \tag{5.55}$$

Although, on the face of it, these quantities appear rather complicated, in reality they merely involve simple arithmetic on quantities trivially computed in a typical GGA calculation.

One subtlety is that GGA calculations are required on both the N- and $(N-1)$-electron systems. The latter is required solely to compute the total electronic energy of the $(N-1)$-electron system, for the computation of I^{GGA}.

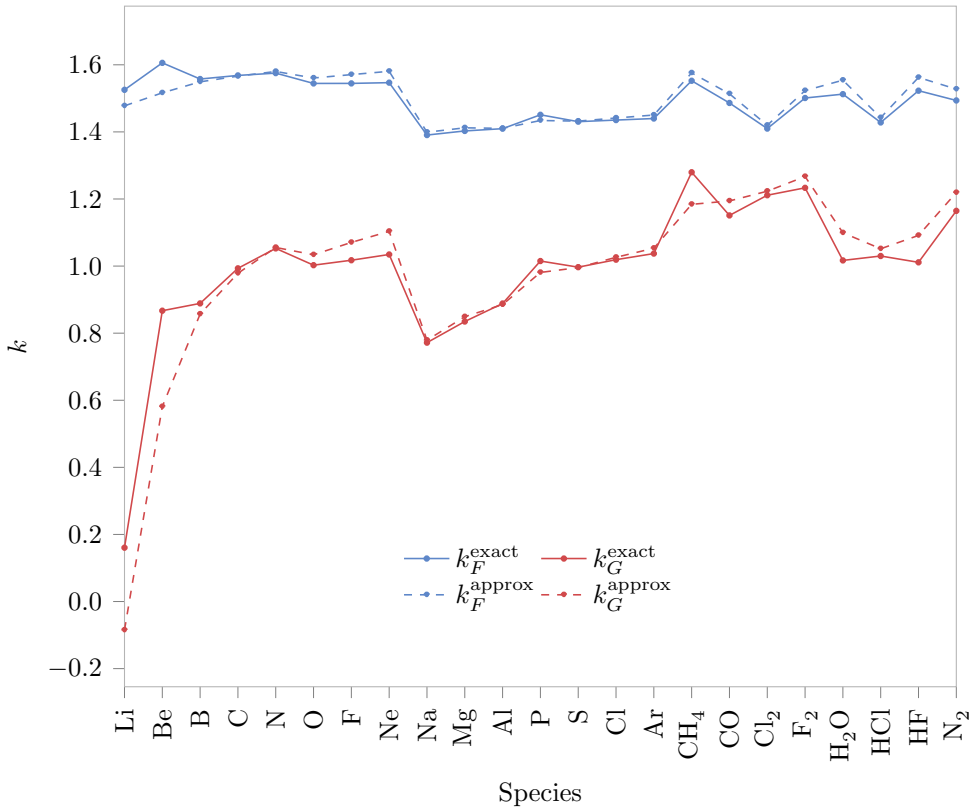

Figure 5.6: Comparison of approximations for k_F and k_G from equations (5.54) and (5.55) with their exact equivalents from Table 5.4.

This need to compute the ionisation potential—whilst arguably inconvenient—is not only found in the current approach but is also central to the conventional functional tuning approaches of previous chapters.

Figure 5.6 plots the near-exact k_F and k_G from Table 5.4 alongside the approximations from equations (5.54) and (5.55), where the approximations use GGA quantities obtained from PBE calculations with the aug-cc-pVTZ basis set. It is clear that the outlined method gives a good approximation of the electron-deficient homogeneity, and so it is reasonable to continue to use the approximate forms in subsequent development.

Given these approximate values of k we could, in principle, repeat the analysis of Figure 5.5 and Table 5.7. Indeed, preliminary investigations revealed very similar optimal values for the prefactors and coordinate-scaling homogeneities However—as already established—the relatively inflexible functional

form associated with a fitted prefactor is insufficient to accurately reproduce exchange–correlation energies. Instead, we argue that since we already require a precursor GGA calculation to compute the approximate, system-dependent k, we can use the same calculation to determine an appropriate system-dependent prefactor.

5.4.4 SYSTEM-DEPENDENT PREFACTOR

Specifically, we demand that the exchange–correlation energy in equations (5.37) and (5.38) equals the GGA value when the GGA density is used, which requires

$$\gamma = \frac{E_{\mathrm{XC}}^{\mathrm{GGA}}}{F^{\mathrm{GGA}}}\,, \tag{5.56}$$

and

$$\alpha = \frac{E_{\mathrm{XC}}^{\mathrm{GGA}} + (J^{\mathrm{GGA}}/N)}{G^{\mathrm{GGA}}}\,. \tag{5.57}$$

Here, F^{GGA} and G^{GGA} are the values of $F[\rho]$ and $G[\rho]$ (we drop the $[\rho]$ for brevity) obtained by evaluating equation (5.12) using the GGA density, for the values of k given by equations (5.54) and (5.55). This can be obtained from a trivial modification of a GGA code.

5.4.5 DETERMINING THE COORDINATE-SCALING PARAMETER

Since we are still working in the realm of a local functional (no gradient dependence, so $n = 0$), the only parameter left to determine is the homogeneity under coordinate scaling, m. Whilst this quantity is again system-dependent, there is no obvious approximation to be made, and so we again search for an optimal value that can be reasonably applied to a variety of systems. With the knowledge of our other parameters we may once again perform a series of self-consistent calculations for a range of m, and hence assign the parameter based on the best results for the given criteria. It must be noted that the value of the prefactor—as determined by the GGA precursor calculation—varies depending on the choice of m, and this must be factored into the subsequent SCF calculations.

For the optimisation criteria, two alternatives immediately spring to mind: minimising the errors in either the exchange–correlation energies or the HOMO

Table 5.8: Optimal values of m based on self-consistent calculations with ED1 and ED2, using a system-dependent prefactor.

Minimisation	Exact k		Approx k	
	ED1	ED2	ED1	ED2
$\lvert\varepsilon_{\mathrm{H}} + I_0\rvert$	0.93	0.98	0.97	1.00

energies. Intuitively, the latter option might be expected to be more useful, as the definition of the prefactors γ and α should ensure that the computed E_{XC} be close to that of the GGA, regardless of m.

Since we are once again fitting a limited subset of systems to find a universal parameter, it makes sense to compare to the best-quality data we have, namely the near-exact ionisation energies (as in Table 5.7), which are appropriate for a functional with $k_{\mathrm{XC}} \approx k_{\mathrm{XC}}^{-}$. That being the case, we must make a decision as to which values of k_F and k_G to use: the near-exact forms of equations (5.39) and (5.43), or the approximate forms of equations (5.54) and (5.55)?

A valid argument could be made for either. The functional, in its final implementation, will use the approximate values by necessity, since exact values are unknown for an arbitrary system, and so it is not unreasonable to use the same values in the initial fit. On the other hand, near-exact values (since we know them for the fitting set) would be consistent with the near-exact quantities to which we are fitting. We consider both options, but ultimately settle on the approximate values, as we feel this to be the most internally consistent procedure.

Table 5.8 lists the optimal values of m, determined to within 0.01, using the $\lvert\varepsilon_{\mathrm{H}} + I_0\rvert$ minimisation criterion for each of the k values. The same set of closed-shell molecules was used, comprising CH_4, CO, Cl_2, F_2, H_2O, HCl, N_2, with a PBE calculation as the GGA precursor and the aug-cc-pVTZ basis set throughout. The precise optimum value varied between molecules, and so m was chosen to minimise the mean absolute error across the set.

It is encouraging that the relevant values are close to unity, given that $F[\rho]$ corresponds directly to the exchange–correlation functional, and $G[\rho]$ to the difference between the exchange–correlation and Fermi–Amaldi functionals: both the Fermi–Amaldi term, and exact exchange (which usually dominates exchange–correlation) are homogeneous of degree one under coordinate scaling.

There is no significant difference between the values derived from near-exact and approximate values of k. This, again, is encouraging, as it reinforces the applicability of the approximate values.

Ultimately, we note that $m = 1$ gives near-optimal HOMO energies and so we use this value throughout. Using a more precise value would ideally require widening the scope of the investigation, and fitting to a much larger set of systems. Moreover, differences in the approximate k may arise though the use of different basis sets (necessary for response calculations), as it makes sense to re-compute k using the same method as the final calculation. This, in turn, may affect the precise value of the optimal m, so we feel that a value of unity is a suitable compromise.

5.5 Summary of the ED functional scheme

Our scheme for a self-consistent calculation on an arbitrary N-electron system can therefore be summarised as follows:

1. Perform GGA calculations on the N- and $(N-1)$-electron systems, and use the data to determine k_F or k_G using equations (5.54) and (5.55).

2. Use the converged GGA density from the above N-electron calculation, together with the calculated k and $m = 1$, to determine the prefactors γ or α from equations (5.56) and (5.57).

3. Finally, perform a self-consistent calculation on the N-electron system using the functional in equations (5.37) and (5.38), with the parameters computed in the above steps.

The remainder of the chapter is devoted to assessing the performance of these two functionals, $E_{\mathrm{XC}}^{\mathrm{ED1}} = \gamma F$, and $E_{\mathrm{XC}}^{\mathrm{ED2}} = -(J/N) + \alpha G$.

5.6 Analysis and performance of the ED functionals

All ED calculations use the PBE functional for the initial GGA calculations; we have confirmed that the results are not sensitive to the choice of GGA.

Table 5.9: Parameters defining the ED1 functional.

Molecule	k_F	γ	$\frac{3k}{3k-m}$	$\frac{3k-m}{3}$
CH_4	1.58	−0.554	1.27	1.24
CO	1.51	−0.593	1.28	1.18
Cl_2	1.42	−0.645	1.31	1.09
F_2	1.52	−0.541	1.28	1.19
H_2O	1.55	−0.592	1.27	1.22
HCl	1.44	−0.649	1.30	1.11
HF	1.56	−0.574	1.27	1.23
N_2	1.53	−0.582	1.28	1.19

Table 5.10: Parameters defining the ED2 functional.

Molecule	k_G	α	$\frac{3k}{3k-m}$	$\frac{3k-m}{3}$
CH_4	1.18	−0.590	1.39	0.85
CO	1.19	−0.696	1.39	0.86
Cl_2	1.22	−0.851	1.37	0.89
F_2	1.27	−0.643	1.36	0.93
H_2O	1.10	−0.639	1.44	0.77
HCl	1.05	−0.913	1.46	0.72
HF	1.09	−0.630	1.44	0.76
N_2	1.22	−0.661	1.38	0.89

The same basis set is used for each of the three stages of the ED calculations. Given that this is a preliminary, proof-of-concept investigation we choose not to compare our results with those from a plethora of other functionals; rather we focus our comparison against PBE, which provides the key ingredients for the ED functionals. We also compare with experimental or near-exact reference values, quoting mean absolute errors (MAEs) relative to these values.

We begin by considering results for eight representative molecules: CH_4, CO, Cl_2, F_2, H_2O, HCl, N_2. Whilst one might argue that these are the same as those used in the 'fitting' of m and hence expected to perform well, we reiterate that a true fit was never undertaken, merely an observation that $m = 1$ gave near-optimal results. Later sections extend the analysis to a wide range of other systems, with different basis sets, and these reinforce the validity of the approach.

Tables 5.9 and 5.10 list the calculated values of k_F and γ for ED1, and k_G and α for ED2, respectively, for each molecule. Also listed are the values of $3k/(3k-m)$ and $(3k-m)/3$, which define the functional forms *via* equation (5.44).

For all eight systems, the density exponent $3k/(3k - m)$ is greater than unity and so the potentials associated with F and G asymptotically vanish, at an exponential rate. It follows that the ED2 exchange–correlation potential in asymptotic regions reduces to the potential of the Fermi–Amaldi functional, which exhibits the exact $-1/r$ form. *The* ED2 *functional therefore yields the exact asymptotic exchange–correlation potential* for each of the eight systems; we have confirmed that this is also the case for all systems considered in this study. By extension, the ED1 functional, in the absence of the Fermi–Amaldi term, exhibits a potential which vanishes too quickly in asymptotic regions.

The effective homogeneities of the ED1 and ED2 functionals are given by equation (5.7), and the central idea behind the functionals is that these values should be close to k_{XC}^{-}. For the eight molecules listed above, the values of k_{XC}^{-} were calculated in Ref. 244, and so we can quantify how well this is achieved in practice.

Figure 5.7 plots the ED1 and ED2 k_{XC} values, obtained by evaluating equation (5.7) using data from self-consistent ED1 and ED2 calculations, explicitly integrating the potential with the converged density and substituting in the final $E_{\mathrm{XC}}[\rho]$. These values are compared with the near-exact k_{XC}^{-} values of Ref. 244. The self-consistent values are very close to the approximate values used to initially define the functionals (deduced from Figure 5.6), and the average discrepancy from near-exact values is less than 2 %. The system-dependence of k_{XC}^{-} is successfully reproduced.

5.6.1 EXCHANGE–CORRELATION ENERGIES

Table 5.11 presents exchange–correlation energies for the initial test set, compared to the near-exact values from Ref. 244. As might be expected, the ED values are very close to those of PBE, with ED2 producing a slightly lower MAE than ED1. Recall the definition of equations (5.56) and (5.57): for the PBE density, the E_{XC} of the ED functionals exactly equals that of PBE, and so the discrepancy between the ED and PBE energies is a measure of the difference between their converged densities. The mean absolute percentage difference between PBE and ED2 exchange–correlation energies is 0.8 %. The discrepancy between total electronic energies (not shown) is just 0.02 %.

For the ED2 functional, the importance of the second term in equation (5.38),

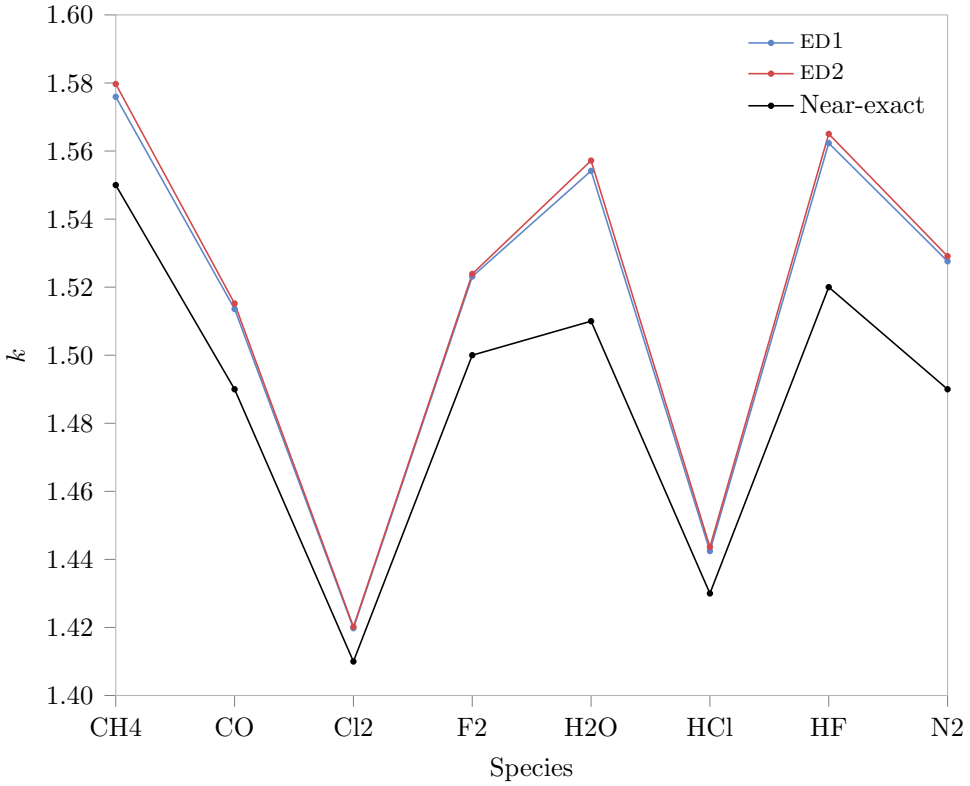

Figure 5.7: Effective homogeneities of the exchange–correlation functional from self-consistent ED1 and ED2 calculations, compared with near-exact k_{XC}^{-} values from Ref. 244.

Table 5.11: Exchange–correlation energies (in E_{h}), compared to near-exact values from Ref. 244.

Molecule	PBE	ED1	ED2	Near-exact
CH_4	−6.836	−6.926	−6.922	−6.865
CO	−13.756	−13.913	−13.835	−13.816
Cl_2	−56.039	−56.297	−56.165	−56.303
F_2	−20.553	−20.813	−20.665	−20.579
H_2O	−9.238	−9.393	−9.373	−9.270
HCl	−28.377	−28.550	−28.491	−28.526
HF	−10.713	−10.901	−10.868	−10.759
N_2	−13.572	−13.727	−13.652	−13.607
MAE	0.080	0.101	0.074	

Table 5.12: HOMO energies (in E_h), compared to the negative of the exact vertical ionisation potential from Ref. 149.

Molecule	PBE	ED1	ED2	$-I^0$
CH_4	−0.347	−0.512	−0.506	−0.526
CO	−0.332	−0.481	−0.515	−0.515
Cl_2	−0.268	−0.344	−0.414	−0.422
F_2	−0.347	−0.554	−0.611	−0.577
H_2O	−0.266	−0.447	−0.466	−0.464
HCl	−0.296	−0.389	−0.425	−0.469
HF	−0.355	−0.574	−0.593	−0.592
N_2	−0.377	−0.551	−0.593	−0.573
MAE	0.194	0.036	0.016	

and its system-dependence, is particularly notable. Removing this term entirely, leaving only the Fermi–Amaldi term, leads to an MAE of 13.007 E_h, illustrating the unsuitability of the unmodified Fermi–Amaldi approximation for computing E_XC. Reinstating the term, but choosing the parameters k_G and α to be an average of the values in Table 5.10, leads to an MAE of 2.501 E_h, which is still two orders of magnitude greater than that of ED2.

5.6.2 HOMO ENERGIES

Table 5.12 presents the HOMO energies for the same test set, compared to $-I_0$, the negative of the exact vertical ionisation potential, from Ref. 149. As is well known, the PBE HOMO energies are significantly above $-I_0$, with an MAE of 0.194 E_h. The ED functionals both show an order of magnitude improvement in the errors, with all HOMO energies being lowered towards $-I_0$. The error of 0.036 E_h for ED1 is almost halved again with ED2, highlighting the beneficial effect of adding the Fermi–Amaldi term.

Again, the importance of the $\alpha G[\rho]$ term in the ED2 functional is highlighted by the MAEs of 0.145 E_h and 0.040 E_h given respectively by Fermi–Amaldi and by averaged parameters. Both errors are again significantly larger than those of ED2.

Table 5.13: HOMO energies (in E_h), compared to the negative of the RMP2 ionisation potential I.

Anion	PBE	ED1	ED2	$-I$
CH_3S^-	0.047	−0.003	−0.060	−0.070
CN^-	−0.001	−0.100	−0.148	−0.146
Cl^-	0.009	−0.046	−0.103	−0.132
F^-	0.057	−0.074	−0.122	−0.134
HOO^-	0.101	−0.006	−0.069	−0.066
NH_2^-	0.094	0.016	−0.028	−0.033
NO_2^-	0.040	−0.065	−0.138	−0.096
OH^-	0.080	−0.019	−0.064	−0.076
PH_2^-	0.056	0.027	−0.028	−0.043
SH^-	0.035	−0.005	−0.061	−0.084
SiH_3^-	0.034	−0.001	−0.043	−0.065
MAE	0.136	0.060	0.016	

5.6.3 BOUND ANIONS

We now proceed to consider the ED functionals for other systems and properties. First, we examine the HOMO energies of bound anions. Most of the molecules in the previous test set do not bind an excess electron and so we consider a different set. Table 5.13 presents HOMO energies for 11 bound anions, determined at MP2/6-31G* anion geometries using the aug-cc-pVTZ basis set. For reference, we list the negative of the vertical ionisation potential of the anion determined using restricted MP2 (RMP2[251]) with the same basis set.

For PBE, the values are essentially all positive and this issue as been the subject of much discussion;[252,253] the MAE is $0.136\,E_h$. With ED1, the HOMO energies are lowered, and all become (correctly) negative with the exception of NH_2^- and PH_2^-. The MAE of $0.060\,E_h$ is not quite as low as for the neutral systems in Table 5.12, but is still less than half that of PBE. For ED2, all the HOMO values become negative, as required, with an MAE of just $0.016\,E_h$; this MAE is the same as that obtained for the neutral systems in Table 5.12.

5.6.4 EXCITATION ENERGIES

Next, we consider TDDFT vertical excitation energies. Table 5.14 presents singlet excitation energies for CO, N_2, and H_2CO, determined at experimental geometries[254] using an augmented Sadlej basis set,[191,255,256] compared to experimental values.[191]

Table 5.14: Vertical excitation energies (in eV), compared to experimental values from Ref. 191. Rydberg and valence excitations are labelled R and V, respectively.

State	Transition	Type	PBE	ED1	ED2	Expt.
CO						
$^1\Sigma^+$	$\sigma \rightarrow 3d\sigma$	R	9.62	13.40	12.33	12.40
$^1\Pi$	$\sigma \rightarrow 3p\pi$	R	9.56	13.32	11.32	11.53
$^1\Sigma^+$	$\sigma \rightarrow 3p\sigma$	R	9.47	13.21	11.59	11.40
$^1\Sigma^+$	$\sigma \rightarrow 3s\sigma$	R	8.99	12.63	10.56	10.78
$^1\Delta$	$\pi \rightarrow \pi^*$	V	10.18	10.52	11.11	10.23
$^1\Sigma^-$	$\pi \rightarrow \pi^*$	V	9.84	10.17	10.39	9.88
$^1\Pi$	$\sigma \rightarrow \pi^*$	V	8.25	8.36	8.04	8.51
	MAE (R)		2.12	1.61	0.17	
	MAE (V)		0.12	0.25	0.62	
	MAE (all)		1.26	1.03	0.36	
N_2						
$^1\Pi_u$	$\pi_u \rightarrow 3s\sigma_g$	R	11.54	15.78	14.52	13.24
$^1\Sigma_u^+$	$\sigma_g \rightarrow 3p\sigma_u$	R	10.47	14.96	13.51	12.98
$^1\Pi_u$	$\sigma_g \rightarrow 3p\pi_u$	R	10.48	15.10	13.29	12.90
$^1\Sigma_g^+$	$\sigma_g \rightarrow 3s\sigma_g$	R	10.23	14.53	12.65	12.20
$^1\Delta_u$	$\pi_u \rightarrow \pi_g$	V	10.08	10.24	10.85	10.27
$^1\Sigma_u^-$	$\pi_u \rightarrow \pi_g$	V	9.66	9.80	9.98	9.92
$^1\Pi_g$	$\sigma_g \rightarrow \pi_g$	V	9.08	9.21	9.01	9.31
	MAE (R)		2.15	2.26	0.66	
	MAE (V)		0.23	0.09	0.32	
	MAE (all)		1.33	1.33	0.51	
H_2CO						
$^1A_2$	$n \rightarrow 3db_1$	R	7.14	11.12	9.64	9.22
$^1A_2$	$n \rightarrow 3pb_1$	R	6.59	10.41	8.46	8.38
$^1B_1$	$\sigma \rightarrow \pi^*$	V	8.85	9.18	8.94	8.68
$^1B_2$	$n \rightarrow 3pa_1$	R	6.38	9.81	8.04	8.12
$^1A_1$	$n \rightarrow 3pb_2$	R	6.40	9.69	8.25	7.97
$^1B_2$	$n \rightarrow 3sa_1$	R	5.73	8.61	7.24	7.09
$^1A_2$	$n \rightarrow \pi^*$	V	3.80	4.00	3.48	3.94
	MAE (R)		1.71	1.77	0.20	
	MAE (V)		0.16	0.28	0.36	
	MAE (all)		1.26	1.34	0.25	

MAEs are presented for Rydberg (R), valence (V), and both categories of excitations combined (all). Accurate Rydberg excitations require[191] the exchange–correlation potential to asymptotically behave as $-1/r + \varepsilon_{\text{H}} + I_0$, which reduces to $-1/r$ when equation (5.1) is exactly satisfied. PBE completely fails to exhibit this form and so the Rydberg excitation energies are much too low, as is well known.

Similarly, the ED1 potential does not exhibit the $-1/r$ asymptotic behaviour, and so the Rydberg excitations remain much too low in energy. Only CO shows any improvement over PBE, and the MAE of 1.610 eV is still very large.

The ED2 potential asymptotically behaves as $-1/r$ and approximately satisfies equation (5.1). As a result, the potential closely resembles the required form and the Rydberg excitations are significantly improved. The improvement is approximately an order of magnitude for CO and H_2CO, but is less pronounced for N_2. This can be traced to the fact that equation (5.1) is less well satisfied for this latter system—the MAEs for the Rydberg excitation energies closely mirror the magnitude of the deviation of $\varepsilon_{\text{H}} + I_0$ from zero.

For the valence excitations in Table 5.14, the performance of ED1 is, surprisingly, better than that of ED2. ED1 shows a small increase in MAE compared to PBE (with the exception of N_2, where the error is significantly lower), whereas ED2 is notably less accurate than PBE in all cases.

5.6.5 Static polarisabilities

Table 5.15 lists static isotropic polarisabilities, determined at experimental geometries[254] using the Sadlej basis set. The table compares the polarisabilities with reference BD(T) values determined using the same basis set.[191] The PBE values are too high, as is well known; both ED1 and ED2 do reduce the values as required, but by significantly too much.

A likely cause of this is that whilst the asymptotic behaviour of the ED2 potential goes correctly as $-1/r$, as desired for Rydberg excitations, the core region of the potential—important for valence excitations and polarisabilities—may still be incorrectly described. The ED1 polarisabilities are even lower, almost doubling the MAE again, suggesting a complete breakdown in the shape of the potential. It is unclear why ED1 valence excitations should show such an improvement over ED2, when the polarisabilities are so much worse, however

Table 5.15: Static isotropic polarisabilities (in a.u.), compared to reference BD(T) values from Ref. 94.

Molecule	PBE	ED1	ED2	Ref.
C_2H_4	28.30	—	24.97	26.91
CH_4	17.40	14.31	15.79	16.43
Cl_2	31.54	27.97	28.70	30.71
CO	13.53	11.06	11.73	13.03
CO_2	17.72	15.02	15.72	17.56
F_2	8.87	7.24	7.68	8.45
H_2O	10.49	7.74	8.53	9.71
H_2S	25.70	21.82	24.18	24.67
HCl	18.26	15.29	16.64	17.43
HF	6.18	4.47	4.86	5.64
N_2	12.13	9.86	10.28	11.75
NH_3	15.37	11.60	12.96	14.33
PH_3	31.85	27.88	31.57	30.44
SO_2	26.44	22.63	24.26	26.06
MAE	0.76	2.25	1.25	

future study into the precise influence of the shape of the potential may offer some insight.

5.6.6 EXCHANGE–CORRELATION POTENTIALS

Figure 5.8 plots the ED1 and ED2 exchange–correlation potentials along the bond axis of two representative systems, CO and PN, compared to both PBE and the near exact $v_{\text{XC}}^{-}(\boldsymbol{r})$ of Ref. 257, determined using the ZMP procedure.[245] The present calculations use the same Huzinaga basis set as was used in Ref. 257. The PBE potentials are well above the near-exact potentials and do not exhibit the desired $-1/r$ asymptotic behaviour; they rapidly decay to zero with increasing distance from the molecule. With ED1, the potentials are lowered towards the near-exact potential (and it is this lowering in energetically important regions that causes ε_{H} to reduce towards $-I_0$), however the asymptotic regions still decay to zero much too quickly. The ED2 potentials are similarly lowered, maintaining the low deviations of ε_{H} from I_0, but crucially they also exhibit the correct $-1/r$ behaviour in the asymptotic regions.

There is, however, clear room for improvement in the non-asymptotic regions. In particular, the ED potentials do not exhibit the intershell structure evident in both PBE and the near-exact potential, and this is a consequence of the fact

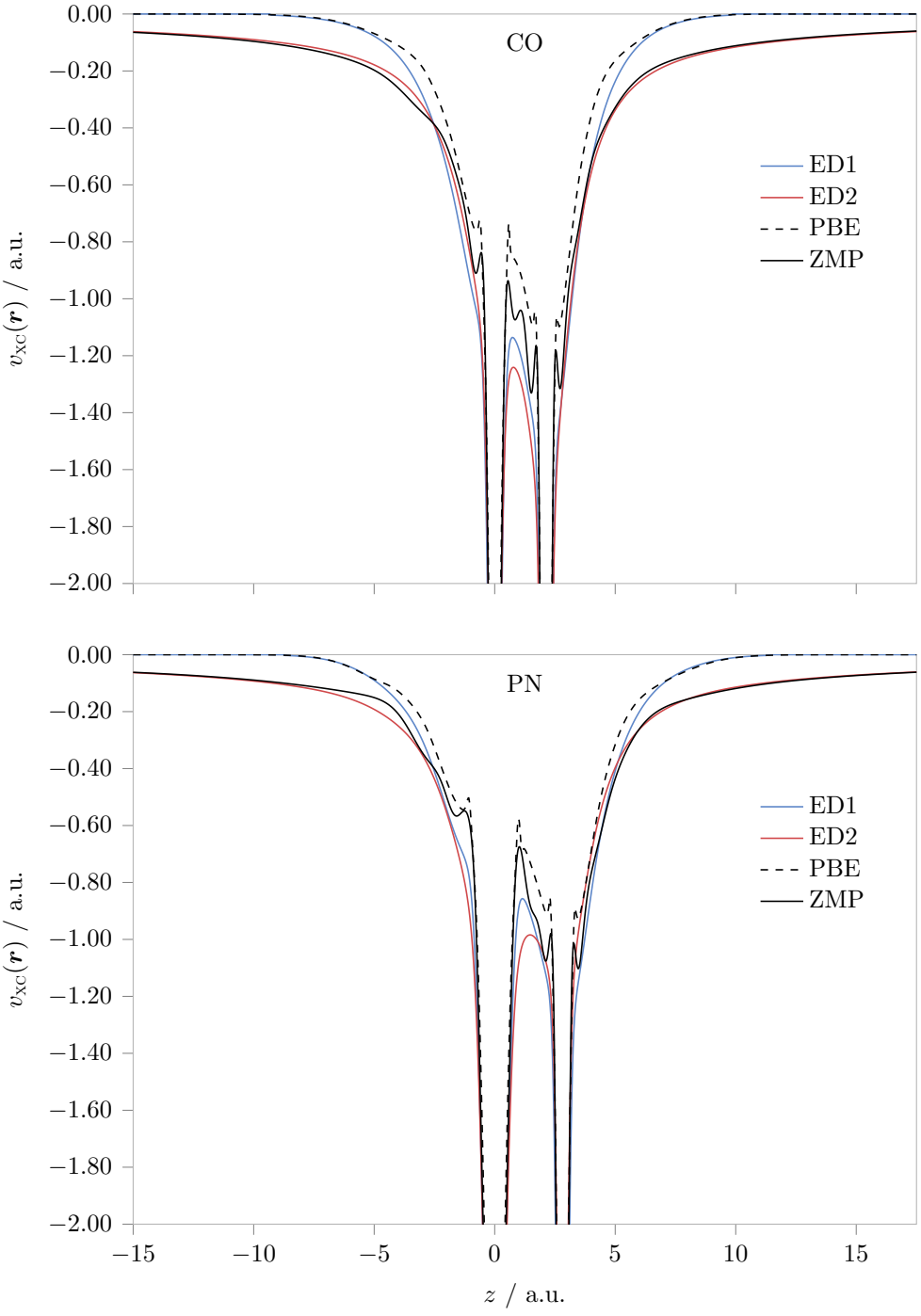

Figure 5.8: Exchange–correlation potentials along the bond axis for CO and PN, compared to the near-exact ZMP potential from Ref. 257.

that $F[\rho]$ and $G[\rho]$ are local, rather than gradient corrected, functionals.

5.6.7 SUMMARY AND EXTENSIONS

Whilst we have continued with the analysis of both ED1 and ED2 throughout, it is clear that in almost all cases (and hence as an overall functional), ED2 is superior, and so we conclude that the addition of the $-J/N$ term is a vital contribution to the success of our functional. We have therefore achieved our original aim: our method yields a functional (ED2) that is appropriate for the electron deficient side of the integer, as illustrated by the effective homogeneities in Figure 5.7, the HOMO energies in Tables 5.12 and 5.13, and the exchange–correlation potentials in Figure 5.8.

The ability to recover the exact asymptotic potential is an added bonus, yielding the improved Rydberg excitation energies in Table 5.14. The ED2 functional is less successful for the valence excitations in Table 5.14, along with the static isotropic polarisabilities in Table 5.15, which is consistent with the lack of quantitative accuracy in the shape of the exchange–correlation potentials in non-asymptotic regions, as shown in Figure 5.8. It is also important to note that the ED functional form is not size-extensive.

We should not be surprised by these deficiencies: the functional form is extremely simple, with a purely local $G[\rho]$ term. Rather, it is encouraging that such improvements can be made with such a simple procedure. The functional development procedure imposes only two system-dependent conditions: we constrain k_{XC} to be close to k_{XC}^{-}, and E_{XC} to be close to the (reasonably accurate) GGA value. It follows from equations (5.7) and (5.9) that the ED2 functional must yield a reasonably accurate $\int v_{\mathrm{XC}}(\boldsymbol{r})\rho(\boldsymbol{r})\,d\boldsymbol{r}$. This is clearly desirable, but is not a sufficient condition to ensure that the potential itself has the correct shape (consistent with the observations in Figure 5.8). Finally we note that the accuracy of the excited states and polarisabilities might also have been affected by neglecting any electric field dependence of the functional parameters.

Looking ahead, an obvious extension of the functional would be to reintroduce gradient dependence into the functional form, *i.e.* to choose $n \neq 0$. This is less trivial than it sounds, however, as without careful choice of the parameters the potential will diverge in asymptotic or zero-density-gradient regions. Some initial investigations were pursued, but SCF convergence problems were quickly

met, coinciding with divergences in the potential. By considering the functional form more closely, we determined that in order to avoid a divergence in the potential, two constraints must be met: $n = \{0\,, 2\,, \geqslant 4\}$, and $m \geqslant kn$.

Given that we know m to be close to unity, and k to be around 1.5, this would appear to limit our use of this functional form to $n = 0$. Furthermore, with increasing n the $G[\rho]$ term quickly becomes unmanageably large, requiring very small values of α to bring it in line with the PBE value, which brings with it its own problems. Further investigation into the nature of the functional form and its derivative must therefore be undertaken before a useful gradient correction can be introduced.

A second extension one might consider would be to use ED2 data to calculate new k and α values and then iterate the approach. Whilst an interesting prospect, care must again be taken because equation (5.55) *relies* on the use of an (incorrect) GGA HOMO energy in order to estimate the derivative discontinuity. If the ED2 HOMO energy was instead used, then the overall effective homogeneity would again be close to 4/3 for all systems.

Thirdly, the dependence of the parameters k_G and α on perturbations, such as electric fields and nuclear coordinates, must be further investigated. Some early investigations were carried out into the stretching of diatomics in order to determine the equilibrium geometry. Although a number of calculations were successful (despite deviations from experimental geometry being an order of magnitude worse than PBE), the majority either failed to converge or encountered other problems. In particular, several diatomics failed to bind at all and for many the character of the cation ground state changed at a particular separation. The latter problem highlights the lack of size-extensivity of the method, given that both the GGA ionisation potential and HOMO energy are vital to the approximation of k.

Finally, regarding the shape of the potential, it may prove fruitful to explicitly enforce the correct shape, in the spirit of Ref. 94.

5.7 Conclusions

In a conceptual departure from the procedures in Chapters 3 and 4 for tuning existing functionals, we have used density scaling considerations as a novel

method of functional development, culminating in an explicit density functional that is appropriate for the *electron-deficient* side of the integer.

By considering a combined density functional $\Omega[\rho]$, comprising both Coulomb and exchange–correlation terms, we determined that multiplying any homogeneous functional (of degree one) by a factor of $(N-1)$ approximately recovers the system-dependent *effective* homogeneity of the exact $\Omega[\rho]$, for a series of atoms. Through careful choice of the homogeneous functional, the procedure was formulated as a correction to the Fermi–Amaldi approximation, such that significant improvements were made in the calculation of the exchange–correlation energy, without losing the correct asymptotic behaviour of the Fermi–Amaldi exchange–correlation potential. The results were further improved by explicitly changing the homogeneity of the corrective functional, although the exact value remained close to unity.

Following the above proof-of-concept procedure, which (approximately) reproduced the *average* effective homogeneity of the $\Omega[\rho]$, and hence the $E_{\mathrm{XC}}[\rho]$ functionals, the same considerations were applied to the *electron-deficient* effective homogeneity. Comparing the Fermi–Amaldi correction (DS: 1) studied by Parr and Ghosh[250] to the near-exact homogeneities calculated by Borgoo *et al.*[244] showed that systems for which the near-exact homogeneity is unity are well described by the Parr–Ghosh functional, and vice versa. To improve on this, we examined a simple, explicit density functional form and imposed on it the near-exact homogeneity from Ref. 244, considering it both as a functional in its own right (ED1), and as a correction to Fermi–Amaldi, following Parr and Ghosh (ED2).

Our initial investigations scaled the functional form by a system-independent prefactor determined by a linear regression fit to known data. Whilst the results were encouraging, producing HOMO energies that were much closer to $-I_0$ than conventional explicit density functionals, the functionals failed to accurately reproduce the exchange–correlation energies, implying that a system-*dependent* prefactor was necessary. In addition, the functional form was explicitly constrained to exhibit the near-exact effective homogeneity which, in the general case, is unknown and must be approximated.

Crucially, the effective homogeneity can be approximated from the results of a precursor GGA calculation, which can also be used to determine the system-

dependent prefactor. This procedure was carried out successfully for both the ED1 and ED2 forms, and the latter in particular showed promising improvements over conventional explicit density functionals, despite maintaining a simple mathematical form.

Compared to GGA results, the ED2 functional yields similar exchange–correlation energies, but HOMO energies that are an order of magnitude closer to the negative of the vertical ionisation potential; for anions, the HOMO energies are negative, as required. Rydberg excitation energies are also notably improved, and the exchange–correlation potential is visibly lowered towards the near-exact potential. Further development is required to improve valence excitations, static isotropic polarisabilities, and the shape of the potential in non-asymptotic regions.

Despite its limitations, the ED2 functional shows promising improvements in several key areas known to be problematic with conventional DFT approaches. Our hope is that the ideas and methods used in its derivation, along with the insight they provide, may prove useful in future study and functional development.

6

Conclusions

For many years, density functional theory has offered chemists a tractable alternative to highly accurate but computationally expensive correlated wave-function methods, for performing electronic structure calculations. Its lower computational cost makes it applicable to much larger, chemically relevant systems, and as a result it has become a widely used and invaluable tool in the field. Given its widespread popularity, much research has been directed towards improving the remaining approximation in what is otherwise an exact theory—the exchange–correlation functional.

Many exchange–correlation functionals exist in the literature, and these have been widely and successfully applied to countless projects. However, many situations remain for which existing functionals fail. This thesis has focused on two particular widespread deficiencies in conventional functionals, which between them are responsible for a large number of common failures. The first, delocalisation error, describes the underestimation of the energy of fractionally charged species, leading to problems such as over-delocalised charge distributions, underestimated reaction barriers, and incorrect orbital energies. The delocalisation error can be conveniently quantified by plotting the variation of the energy E with the number of electrons N, for both fractional and integer N (an E *vs* N plot). The exact behaviour should be a piecewise-linear variation between pairs of integer N, however conventional DFT functionals are

convex at fractional N. The second deficiency concerns the incorrect long-range asymptotic behaviour of the exchange–correlation potential, leading to, for example, underestimated Rydberg excitation energies. We considered some novel approaches to correcting these deficiencies, and suggested avenues for future work.

Chapter 3 examined the balance between the short-range DFT exchange and long-range exact exchange components of range-separated hybrid (RSH) functionals, and its effect on the delocalisation error. By tuning the range-separation parameter μ (*i.e.* the rate at which exact exchange is incorporated) in a *system-dependent* manner, in order to achieve near-linearity in E *vs* N, it has previously been shown that much more accurate frontier orbital energies and energy differences can be obtained. This effect can be seen as resulting from a reduction in the delocalisation error, due to a cancellation between the delocalisation of DFT exchange, and the *localisation* of Hartree–Fock (exact) exchange.

The chapter began by examining the justification for this tuning procedure, and the effect of varying μ on successively ionised species. Most interestingly, the optimal μ varied considerably for each successive species, suggesting that achieving near-linearity in the whole E *vs* N plot is not feasible with a single value of μ. Since the most successful tuning procedures rely on attempting to constrain two adjacent E *vs* N segments simultaneously, a systematic assessment was subsequently performed of the tuning procedures, relating its successes and failures to the explicit E *vs* N behaviour. It was demonstrated that, for the double-segment tuning procedure, the optimal μ returned E *vs* N plots with some residual non-linearity in both sections. However, the errors due to this non-linearity conveniently cancelled with errors in the integer-N energy differences, to give the observed low errors in frontier orbital energy differences. Whilst not a rigorous or universally applicable solution to the delocalisation error, this marks an important step in successfully computing quantities affected by the delocalisation error, and in understanding some of its causes.

Following the successful tuning of RSH functionals to give near-linearity in the *energy* variation with fractional N, Chapter 4 considered whether the same functionals are similarly optimal for the *density* variation, which should

also be piecewise-linear. The Fukui function in conceptual DFT, defined as the derivative of the density with respect to N, is typically approximated as the difference between the densities of adjacent integer-N species. By defining an alternate finite difference method, the density difference between the integer species and a small fractional N, we identified *initial* Fukui functions, analogous to the initial slopes (*i.e.* frontier orbital energies) of the E *vs* N plots. Comparison of initial Fukui functions to those from integer species, computed using optimally tuned functionals from Chapter 3, indicated that the optimal μ for near-linearity in the energy variation with N can differ significantly from the optimal μ for density variation. Although only one test system was considered, the results suggested that a system-independent value of μ optimised for thermochemical properties may be most appropriate for the Fukui function.

A second quantity important to conceptual DFT, the electronegativity, was then considered. Since this quantity can be calculated using the frontier orbital energies, it was hypothesised that functionals optimally tuned for the energy should provide a better approximation than their non-tuned equivalents (unlike for the Fukui function). This was indeed found to be the case, reinforcing the conclusion that one cannot find even a system-dependent μ that is universally appropriate for all computed properties.

Whilst the tuning procedure provides an accessible approach to the successful calculation of problematic quantities, its main drawback is that for each new property and system, many calculations have to be performed to determine the optimal μ. This process is not only time consuming and labour-intensive, it also severely limits the transferability of a particular set of functional parameters to a variety of different problems. Furthermore, there is currently no systematic means of verifying that the convenient error cancellation will hold for a particular system or property.

As such, a more important goal is to find a more universally applicable functional that is inherently free from the delocalisation error. Chapter 5 presented an entirely new approach to functional development, based on homogeneity properties and known scaling relations, culminating in an *electron-deficient* (ED) explicit density functional that both satisfies the exact Koopmans condition $\varepsilon_{\text{H}} = -I_0$ associated with energy-linearity, and exhibits a potential with the

correct long-range asymptotic behaviour.

After establishing that a simple functional form with known homogeneity properties could be successfully used to approximately reproduce near-exact average effective homogeneities and exchange–correlation energies, it was demonstrated that similar homogeneity arguments could be used to generate a functional with a potential associated with the electron-deficient side of an integer species. In other words, the potential would be appropriate for satisfying $\varepsilon_{\text{H}} = -I$. Due to the highly system-dependent nature of the exact electron-deficient effective homogeneity, this quantity was approximated, along with a scaling prefactor, by a precursor GGA calculation; the resulting functional was termed ED1. Alternatively, by treating the functional as a correction to the Fermi–Amaldi approximation, the correct asymptotic behaviour of the potential could also be recovered, leading to a second functional, denoted ED2. The functionals are unlike any other functionals proposed previously. Although the functional form is still inherently system-dependent, the precursor calculations to determine the system-dependent parameters are much easier and less time-consuming than the process in Chapter 3.

The two functionals were implemented self-consistently, and assessed in terms of their ability to reproduce exchange–correlation energies, HOMO energies of neutral molecules and bound anions, excitation energies and static polarisabilities. The ED2 functional in particular showed great promise despite such a simple mathematical form. Exchange–correlation energies were successfully reproduced, whilst HOMO energies for both neutral molecules and bound anions were considerably closer to $-I_0$ than for PBE. Rydberg excitation energies were also much improved due to the correct asymptotic behaviour of the potential.

Several challenges still remain. In particular, large errors in the valence excitation energies and static polarisabilities indicate improvements to be made in the shape of the potential in core regions. Secondly, the approximation relies on precursor GGA calculations on both the system in question and its cation: eliminating this step would cut down on both the computation time and complexity of the method. Further investigation is also needed into improving the ED functionals through introduction of a gradient correction; initial investigations were carried out, but many natural choices for the parameters introduced

divergences into the potential. Finally, the applicability of these functionals to other properties must be addressed: since they are explicitly constructed to improve the HOMO energy and (in the case of ED2) electron-deficient potential, their general applicability may be limited. In particular, the methods are not size-extensive, and calculations such as geometry optimisations require iterative determination of the functional parameters at each step (or a knowledge of their dependence on perturbations such as varying nuclear coordinates).

Overall, the fact that a local, explicit density functional can offer such a great improvement over conventional alternatives is very encouraging. Our hope is that the insights into the nature of linearity in the energy and density variation with respect to N of Chapters 3 and 4, coupled with the benefits of satisfying correct homogeneity conditions highlighted in Chapter 5, will aid and inspire future functional development.

A

ED Exchange–correlation kernel

Consider the general functional

$$E[\rho] = \left[\int \rho(\boldsymbol{r})^p \, d\boldsymbol{r} \right]^q = \varepsilon[\rho]^q \,. \tag{A.1}$$

We define the functional derivative by consideration of an infinitesimal change in the density $\delta\rho(\boldsymbol{r})$, and expanding as a Taylor series,

$$\begin{aligned} E[\rho + \delta\rho] &= E[\rho] + \int \frac{\delta E[\rho]}{\delta\rho(\boldsymbol{r})} \delta\rho(\boldsymbol{r}) \, d\boldsymbol{r} \\ &\quad + \frac{1}{2} \iint \frac{\delta^2 E[\rho]}{\delta\rho(\boldsymbol{r})\delta\rho(\boldsymbol{r}')} \delta\rho(\boldsymbol{r})\delta\rho(\boldsymbol{r}') \, d\boldsymbol{r} \, d\boldsymbol{r}' + \cdots \,. \\ &= \varepsilon[\rho + \delta\rho]^q \end{aligned} \tag{A.2}$$

Similarly, we can expand the "inner" functional $\varepsilon[\rho]$ as

$$\begin{aligned} \varepsilon[\rho + \delta\rho] = \varepsilon[\rho] + \Bigg\{ &\int \frac{\delta\varepsilon[\rho]}{\delta\rho(\boldsymbol{r})} \delta\rho(\boldsymbol{r}) \, d\boldsymbol{r} \\ &+ \frac{1}{2} \iint \frac{\delta^2\varepsilon[\rho]}{\delta\rho(\boldsymbol{r})\delta\rho(\boldsymbol{r}')} \delta\rho(\boldsymbol{r})\delta\rho(\boldsymbol{r}') \, d\boldsymbol{r} \, d\boldsymbol{r}' + \cdots \Bigg\} \,, \end{aligned} \tag{A.3}$$

and so if we collect everything in the brackets into a term denoted "B", we can represent $\varepsilon[\rho+\delta\rho]^q$ as a binomial expansion

$$\varepsilon[\rho+\delta\rho]^q = \varepsilon[\rho]^q + q\varepsilon[\rho]^{q-1}\{B\} + \frac{1}{2}q(q-1)\varepsilon[\rho]^{q-2}\{B\}^2 + \cdots . \tag{A.4}$$

Thus,

$$\begin{aligned} E[\rho+\delta\rho] - E[\rho] &= \varepsilon[\rho+\delta\rho]^q - \varepsilon[\rho]^q \\ &= q\varepsilon[\rho]^{q-1}\left\{\int \frac{\delta\varepsilon[\rho]}{\delta\rho(\boldsymbol{r})}\delta\rho(\boldsymbol{r})\,d\boldsymbol{r} + \frac{1}{2}\iint \frac{\delta^2\varepsilon[\rho]}{\delta\rho(\boldsymbol{r})\delta\rho(\boldsymbol{r}')}\delta\rho(\boldsymbol{r})\delta\rho(\boldsymbol{r}')\,d\boldsymbol{r}\,d\boldsymbol{r}'\right\} \\ &\quad + \frac{1}{2}\,q(q-1)\varepsilon[\rho]^{q-2}\left\{\int \frac{\delta\varepsilon[\rho]}{\delta\rho(\boldsymbol{r})}\delta\rho(\boldsymbol{r})\,d\boldsymbol{r}\int \frac{\delta\varepsilon[\rho]}{\delta\rho(\boldsymbol{r}')}\delta\rho(\boldsymbol{r}')\,d\boldsymbol{r}'\right\}, \end{aligned} \tag{A.5}$$

where we have truncated the expansions at the quadratic level.

By collecting terms of each power, we can therefore identify the first and second functional derivatives of the original energy functional $E[\rho]$,

$$\frac{\delta E[\rho]}{\delta\rho(\boldsymbol{r})} = q\varepsilon[\rho]^{q-1}\,\frac{\delta\varepsilon[\rho]}{\delta\rho(\boldsymbol{r})}\,, \tag{A.6}$$

and

$$\begin{aligned} \frac{\delta^2 E[\rho]}{\delta\rho(\boldsymbol{r})\delta\rho(\boldsymbol{r}')} &= q\varepsilon[\rho]^{q-1}\,\frac{\delta^2\varepsilon[\rho]}{\delta\rho(\boldsymbol{r})\delta\rho(\boldsymbol{r}')} \\ &\quad + q(q-1)\varepsilon[\rho]^{q-2}\,\frac{\delta\varepsilon[\rho]}{\delta\rho(\boldsymbol{r})}\frac{\delta\varepsilon[\rho]}{\delta\rho(\boldsymbol{r}')}\,. \end{aligned} \tag{A.7}$$

We now consider the derivatives of $\varepsilon[\rho]$. From the definitions

$$\varepsilon[\rho] = \int \rho(\boldsymbol{r})^p\,d\boldsymbol{r} \tag{A.8}$$

and

$$\begin{aligned} \varepsilon[\rho+\delta\rho] &= \int \big\{\rho(\boldsymbol{r}) + \delta\rho(\boldsymbol{r})\big\}^p\,d\boldsymbol{r} \\ &= \int \rho(\boldsymbol{r})^p + \int p\rho(\boldsymbol{r})^{p-1}\,\delta\rho(\boldsymbol{r})\,d\boldsymbol{r} \\ &\quad + \frac{1}{2}\int p(p-1)\rho(\boldsymbol{r})^{p-2}\,\delta\rho(\boldsymbol{r})^2\,d\boldsymbol{r} + \cdots \end{aligned} \tag{A.9}$$

we can, from the above expansions, identify

$$\frac{\delta\varepsilon[\rho]}{\delta\rho(\boldsymbol{r})} = p\rho(\boldsymbol{r})^{p-1}\,, \tag{A.10}$$

and

$$\frac{\delta^2\varepsilon[\rho]}{\delta\rho(\boldsymbol{r})\delta\rho(\boldsymbol{r}')} = p(p-1)\rho(\boldsymbol{r})^{p-2}\,\delta(\boldsymbol{r}-\boldsymbol{r}')\,, \tag{A.11}$$

where the introduction of the delta functional collapses $\boldsymbol{r}'$ to $\boldsymbol{r}$ and reduces the double integral of the quadratic term in equation (A.3) to the single integral of equation (A.9).

Substituting equation (A.10) into equation (A.6) recovers the potential, equation (5.48)

$$v_E = \frac{\delta E[\rho]}{\delta\rho(\boldsymbol{r})} = pq\varepsilon[\rho]^{q-1}\,\rho(\boldsymbol{r})^{\,p-1}\,. \tag{A.12}$$

Substituting equation (A.11) into the first term of equation (A.7), and equation (A.10) *twice* into the second term (maintaining the $\boldsymbol{r}$- and $\boldsymbol{r}'$-dependence), yields our component of the exchange–correlation kernel, equation (5.49)

$$\begin{aligned}\frac{\delta^2 E[\rho]}{\delta\rho(\boldsymbol{r})\delta\rho(\boldsymbol{r}')} &= qp(p-1)\varepsilon[\rho]^{q-1}\,\rho(\boldsymbol{r})^{p-2}\,\delta(\boldsymbol{r}-\boldsymbol{r}') \\ &\quad + p^2q(q-1)\varepsilon[\rho]^{q-2}\,\rho(\boldsymbol{r})^{p-1}\,\rho(\boldsymbol{r}')^{p-1}\,.\end{aligned} \tag{A.13}$$

In order to evaluate the two-electron integrals of the kernel, we first note that

$$(ia|f_{\mathrm{XC}}^{\mathrm{ED}}|jb) = \iint d\boldsymbol{r}\,d\boldsymbol{r}'\,\frac{\delta^2E[\rho]}{\delta\rho(\boldsymbol{r})\delta\rho(\boldsymbol{r}')}\,\varphi_i(\boldsymbol{r})\varphi_a(\boldsymbol{r})\varphi_j(\boldsymbol{r}')\varphi_b(\boldsymbol{r}') \tag{A.14}$$

$$\begin{aligned}&= \iint d\boldsymbol{r}\,d\boldsymbol{r}'\,\Big\{qp(p-1)\varepsilon[\rho]^{q-1}\rho(\boldsymbol{r})^{p-2}\,\delta(\boldsymbol{r}-\boldsymbol{r}') \\ &\quad + p^2q(q-1)\varepsilon[\rho]^{q-2}\rho(\boldsymbol{r})^{p-1}\rho(\boldsymbol{r}')^{p-1}\Big\}\,\varphi_i(\boldsymbol{r})\varphi_a(\boldsymbol{r})\varphi_j(\boldsymbol{r}')\varphi_b(\boldsymbol{r}')\,.\end{aligned} \tag{A.15}$$

The first term reduces to a single integral in terms of $\boldsymbol{r}$, due to the delta function, whereas the second term decomposes into separate integrals involving the respective orbitals,

$$(ia|f_{\mathrm{XC}}^{\mathrm{ED}}|jb) = \iint d\boldsymbol{r}\,d\boldsymbol{r}'\,\frac{\delta^2E[\rho]}{\delta\rho(\boldsymbol{r})\delta\rho(\boldsymbol{r}')}\,\varphi_i(\boldsymbol{r})\varphi_a(\boldsymbol{r})\varphi_j(\boldsymbol{r}')\varphi_b(\boldsymbol{r}') \tag{A.16}$$

$$\begin{aligned}&= qp(p-1)\varepsilon[\rho]^{q-1}\int d\boldsymbol{r}\,\rho(\boldsymbol{r})^{p-2}\,\varphi_i(\boldsymbol{r})\varphi_a(\boldsymbol{r})\varphi_j(\boldsymbol{r})\varphi_b(\boldsymbol{r}) \\ &\quad + p^2q(q-1)\varepsilon[\rho]^{q-2}\int d\boldsymbol{r}\,\rho(\boldsymbol{r})^{p-1}\varphi_i(\boldsymbol{r})\varphi_a(\boldsymbol{r})\int d\boldsymbol{r}'\,\rho(\boldsymbol{r}')^{p-1}\,\varphi_j(\boldsymbol{r}')\varphi_b(\boldsymbol{r}')\,,\end{aligned} \tag{A.17}$$

to give us the final expression in equation (5.50).

B

Presented work

Publications

A significant portion of the work in this thesis has been published (or is due to be submitted) in peer-reviewed journals:

1. Assessment of tuning methods for enforcing approximate energy linearity in range–separated hybrid functionals
   J. D. Gledhill, M. J. G. Peach and D. J. Tozer, *J. Chem. Theory Comput.* **9**, 4414–4418 (2013).

2. System-dependent exchange–correlation functional with exact asymptotic potential and $\varepsilon_{\text{H}} = -I$
   J. D. Gledhill and D. J. Tozer, *J. Chem. Phys.* **143**, 024101 (2015).

3. Tuning parameter in range-separated exchange–correlation functionals: successive ionisations and the Fukui function
   J. D. Gledhill, F. De Proft and D. J. Tozer, *in prep.*

Meetings attended

I have presented my work at a number of schools, meetings and symposia, including:

1. National Training School in Theoretical Chemistry
   2012, Oxford
   Poster prize winner: *Tuned range–separated hybrid functionals in DFT.*

2. Chemistry Postgraduate Gala Symposium
   2013, Durham
   Poster: *Tuned range–separated hybrid functionals in DFT.*

3. International Conference on Density Functional Theory and its Applications
   2013, Durham
   Poster: *Energy linearity in tuned range–separated hybrid functionals.*

4. Chemistry Postgraduate Gala Symposium
   2014, Durham
   Talk: *Novel approaches to improving the exchange–correlation functional in DFT.*

5. European Seminar on Computational Methods in Quantum Chemistry
   2014, Houffalize
   Poster: *Assessment of tuning methods for enforcing approximate energy linearity in range–separated hybrid functionals.*

Bibliography

1. P. A. M. Dirac, *The Principles of Quantum Mechanics*, Oxford University Press, 4th edn., 1958.
2. A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Macmillan Publishing Company, New York, 1982.
3. P. Atkins and R. Friedman, *Molecular Quantum Mechanics*, Oxford University Press, 5th edn., 2011.
4. M. Planck, *Ann. Phys.*, 1901, **4**, 553–563.
5. A. Einstein, *Ann. Phys.*, 1905, **17**, 132–148.
6. N. Bohr, *Phil. Mag.*, 1913, **6**, 1–25.
7. L. de Broglie, *Ann. Phys.*, 1925, **3**, 22–128.
8. E. Schrödinger, *Ann. Phys.*, 1926, **79**, 361–376.
9. E. Schrödinger, *Ann. Phys.*, 1926, **79**, 489–527.
10. E. Schrödinger, *Naturwissenschaften*, 1926, **28**, 664–666.
11. E. Schrödinger, *Ann. Phys.*, 1926, **79**, 734–756.
12. E. Schrödinger, *Ann. Phys.*, 1926, **80**, 437–490.
13. E. Schrödinger, *Ann. Phys.*, 1926, **81**, 109–139.
14. W. Heisenberg, *Z. Phys.*, 1925, **33**, 879–893.

15. M. Born, *Z. Phys.*, 1926, **38**, 803–827.

16. M. Born and J. R. Oppenheimer, *Ann. Phys.*, 1927, **84**, 457–484.

17. W. Pauli, *Z. Phys.*, 1927, **43**, 601–623.

18. J. C. Slater, *Phys. Rev.*, 1929, **34**, 1293–1322.

19. J. C. Slater, *Phys. Rev.*, 1930, **35**, 509–529.

20. D. R. Hartree, *Proc. Camb. Phil. Soc.*, 1928, **24**, 89–110.

21. D. R. Hartree, *Proc. Camb. Phil. Soc.*, 1928, **24**, 111–132.

22. D. R. Hartree, *Proc. Camb. Phil. Soc.*, 1928, **24**, 426–437.

23. V. Fock, *Z. Phys.*, 1930, **61**, 126–148.

24. V. Fock, *Z. Phys.*, 1930, **62**, 795–805.

25. C. C. J. Roothaan, *Rev. Mod. Phys.*, 1951, **23**, 69–89.

26. G. G. Hall, *Proc. R. Soc. Lond. A*, 1951, **205**, 541–552.

27. J. C. Slater, *Phys. Rev.*, 1930, **36**, 57–64.

28. C. Zener, *Phys. Rev.*, 1930, **36**, 51–56.

29. S. F. Boys, *Proc. R. Soc. Lond. A*, 1950, **200**, 542–554.

30. J. A. Pople and R. K. Nesbet, *J. Chem. Phys.*, 1954, **22**, 571–572.

31. P.-O. Löwdin, in *Advances in Chemical Physics, Volume 2*, Interscience Pubishers, 1959, pp. 207–322.

32. L. Brillouin, *Actualities Sci. Ind*, 1934, **71**, 159.

33. S. R. Langhoff and E. R. Davidson, *Int. J. Quantum Chem.*, 1974, **8**, 61–72.

34. J. Čížek, *J. Chem. Phys.*, 1966, **45**, 4256–4266.

35. J. Čížek and J. Paldus, *Int. J. Quantum Chem.*, 1971, **5**, 359–379.

36. J. Čížek, *Theor. Chim. Acta.*, 1991, **80**, 91–94.

37. J. Noga and R. J. Bartlett, *J. Chem. Phys.*, 1987, **86**, 7041–7050.

38. O. Christiansen, H. Koch and P. Jørgensen, *Chem. Phys. Lett.*, 1995, **243**, 409–418.

39. O. Christiansen, *Theor. Chem. Acc.*, 2006, **116**, 106–123.

40. O. Christiansen, H. Koch and P. Jørgensen, *J. Chem. Phys.*, 1995, **103**, 7429–7441.

41. H. Koch, O. Christiansen, P. Jørgensen, A. M. de Meras and T. Helgaker, *J. Chem. Phys.*, 1997, **106**, 1808–1818.

42. K. A. Brueckner, *Phys. Rev.*, 1954, **96**, 508–516.

43. R. K. Nesbet, *Phys. Rev.*, 1958, **109**, 1632–1638.

44. N. C. Handy, J. A. Pople, M. Head-Gordon, K. Raghavachari and G. W. Trucks, *Chem. Phys. Lett.*, 1989, **164**, 185–192.

45. C. Møller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 618–622.

46. K. Raghavachari, G. W. Trucks, J. A. Pople and M. Head-Gordon, *Chem. Phys. Lett.*, 1989, **157**, 479–483.

47. L. H. Thomas, *Proc. Camb. Phil. Soc.*, 1927, **23**, 542–548.

48. E. Fermi, *Z. Phys.*, 1928, **48**, 73–79.

49. P. A. M. Dirac, *Proc. Cam. Phil. Soc.*, 1930, **26**, 376–385.

50. P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **136**, B864–B871.

51. R. G. Parr, R. A. Donnelly, M. Levy and W. E. Palke, *J. Chem. Phys.*, 1978, **68**, 3801–3807.

52. W. Kohn, A. D. Becke and R. G. Parr, *J. Phys. Chem.*, 1996, **100**, 12974–12980.

53. P. Geerlings, F. De Proft and W. Langenaeker, *Chem. Rev.*, 2003, **103**, 1793–1874.

54. E. H. Lieb, *Int. J. Quantum Chem.*, 1983, **24**, 243–277.

55. M. Levy, *Proc. Natl. Acad. Sci. USA*, 1979, **76**, 6062–6065.

56. M. Levy, *Phys. Rev. A*, 1982, **26**, 1200–1208.

57. W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133–A1138.

58. R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, 1989, pp. 142–200.

59. D. M. Ceperley and B. J. Alder, *Phys. Rev. Lett.*, 1980, **45**, 566–569.

60. S. H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, 1980, **58**, 1200–1211.

61. J. P. Perdew and Y. Wang, *Phys. Rev. B*, 1992, **45**, 13244–13249.

62. J. P. Perdew and W. Yue, *Phys. Rev. B*, 1986, **33**, 8800–8802.

63. J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.

64. A. D. Becke, *J. Chem. Phys.*, 1986, **84**, 4524–4529.

65. A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.

66. C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.

67. R. Colle and O. Salvetti, *Theor. Chim. Acta.*, 1975, **37**, 329–334.

68. A. D. Becke, *J. Chem. Phys.*, 1992, **97**, 9173–9177.

69. J. P. Perdew, S. Kurth, A. Zupan and P. Blaha, *Phys. Rev. Lett.*, 1999, **82**, 2544–2547.

70. J. Tao, J. P. Perdew, V. N. Staroverov and G. E. Scuseria, *Phys. Rev. Lett.*, 2003, **91**, 146401.

71. J. P. Perdew, J. Tao, V. N. Staroverov and G. E. Scuseria, *J. Chem. Phys.*, 2004, **120**, 6898–6911.

72. T. Van Voorhis and G. E. Scuseria, *J. Chem. Phys.*, 1998, **109**, 400.

73. Y. Zhao and D. G. Truhlar, *J. Chem. Phys.*, 2006, **125**, 194101.

74. Y. Zhao and D. G. Truhlar, *Acc. Chem. Res.*, 2008, **41**, 157–167.

75. R. Peverati and D. G. Truhlar, *J. Phys. Chem. Lett.*, 2012, **3**, 117–124.

76. A. D. Boese and J. M. L. Martin, *J. Chem. Phys.*, 2004, **121**, 3405.

77. Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241.

78. R. Peverati and D. G. Truhlar, *J. Phys. Chem. Lett.*, 2011, **2**, 2810–2817.

79. A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 1372–1377.

80. A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

81. P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.

82. A. Seidl, A. Görling, P. Vogl, J. Majewski and M. Levy, *Phys. Rev. B*, 1996, **53**, 3764–3774.

83. A. D. Becke, *J. Chem. Phys.*, 2005, **122**, 64101.

84. E. R. Johnson and A. D. Becke, *J. Chem. Phys.*, 2005, **123**, 024101.

85. E. R. Johnson and A. D. Becke, *J. Chem. Phys.*, 2006, **124**, 174104.

86. A. D. Becke and E. R. Johnson, *J. Chem. Phys.*, 2007, **127**, 124108.

87. E. R. Johnson and A. D. Becke, *J. Chem. Phys.*, 2008, **128**, 124105.

88. M. Ernzerhof and G. E. Scuseria, *J. Chem. Phys.*, 1999, **110**, 5029–5036.

89. C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158–6170.

90. A. D. Becke, *J. Chem. Phys.*, 1996, **104**, 1040–1046.

91. J. P. Perdew, M. Ernzerhof and K. Burke, *J. Chem. Phys.*, 1996, **105**, 9982–9985.

92. A. D. Becke, *J. Chem. Phys.*, 1997, **107**, 8554–8560.

93. F. A. Hamprecht, A. J. Cohen, D. J. Tozer and N. C. Handy, *J. Chem. Phys.*, 1998, **109**, 6264–6271.

94. P. J. Wilson, T. J. Bradley and D. J. Tozer, *J. Chem. Phys.*, 2001, **115**, 9233–9242.

95. T. W. Keal and D. J. Tozer, *J. Chem. Phys.*, 2005, **123**, 121103.

96. P. Mori-Sánchez, A. J. Cohen and W. Yang, *J. Chem. Phys.*, 2006, **124**, 091102.

97. A. J. Cohen, P. Mori-Sánchez and W. Yang, *J. Chem. Phys.*, 2007, **126**, 191109.

98. A. J. Cohen, P. Mori-Sánchez and W. Yang, *J. Chem. Phys.*, 2007, **127**, 034101.

99. M. Ernzerhof, *Chem. Phys. Lett.*, 1996, **263**, 499–506.

100. T. Leininger, H. Stoll, H.-J. Werner and A. Savin, *Chem. Phys. Lett.*, 1997, **275**, 151–160.

101. R. D. Adamson, J. P. Dombroski and P. M. W. Gill, *J. Comput. Chem.*, 1999, **20**, 921–927.

102. P. M. W. Gill, R. D. Adamson and J. A. Pople, *Mol. Phys.*, 1996, **88**, 1005–1010.

103. A. Savin, in *Recent Developments and Applications of Modern Density Functional Theory*, ed. J. M. Seminario, Elsevier, Amsterdam, 1996, pp. 327–357.

104. H. Iikura, T. Tsuneda, T. Yanai and K. Hirao, *J. Chem. Phys.*, 2001, **115**, 3540–3544.

105. M. Kamiya, T. Tsuneda and K. Hirao, *J. Chem. Phys.*, 2002, **117**, 6010.

106. Y. Tawada, T. Tsuneda, S. Yanagisawa, T. Yanai and K. Hirao, *J. Chem. Phys.*, 2004, **120**, 8425–8433.

107. T. Sato, T. Tsuneda and K. Hirao, *Mol. Phys.*, 2005, **103**, 1151–1164.

108. T. Sato, T. Tsuneda and K. Hirao, *J. Chem. Phys.*, 2005, **123**, 104307.

109. M. Kamiya, H. Sekino, T. Tsuneda and K. Hirao, *J. Chem. Phys.*, 2005, **122**, 234111.

110. M. Chiba, T. Tsuneda and K. Hirao, *J. Chem. Phys.*, 2006, **124**, 144106.

111. T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51–57.

112. J. Toulouse, F. Colonna and A. Savin, *Phys. Rev. A*, 2004, **70**, 062505.

113. J. Heyd, G. E. Scuseria and M. Ernzerhof, *J. Chem. Phys.*, 2003, **118**, 8207–8215.

114. J. Heyd and G. E. Scuseria, *J. Chem. Phys.*, 2004, **120**, 7274–7280.

115. J. Heyd and G. E. Scuseria, *J. Chem. Phys.*, 2004, **121**, 1187–1192.

116. J. Heyd, J. E. Peralta, G. E. Scuseria and R. L. Martin, *J. Chem. Phys.*, 2005, **123**, 174101.

117. E. N. Brothers, A. F. Izmaylov, J. O. Normand, V. Barone and G. E. Scuseria, *J. Chem. Phys.*, 2008, **129**, 011102.

118. O. A. Vydrov, J. Heyd, A. V. Krukau and G. E. Scuseria, *J. Chem. Phys.*, 2006, **125**, 074106.

119. E. Runge and E. K. U. Gross, *Phys. Rev. Lett.*, 1984, **52**, 997–1000.

120. A. Dreuw and M. Head-Gordon, *Chem. Rev.*, 2005, **105**, 4009–4037.

121. E. K. U. Gross, C. A. Ullrich and U. J. Gossmann, in *Density Functional Theory*, ed. E. K. U. Gross and R. M. Dreizler, Plenum, New York, 1994, pp. 149–171.

122. M. A. L. Marques and E. K. U. Gross, *Annu. Rev. Phys. Chem.*, 2004, **55**, 427–455.

123. M. A. L. Marques and E. K. U. Gross, in *Time-Dependent Density Functional Theory*, 2003, pp. 144–184.

124. A. J. Cohen, P. Mori-Sánchez and W. Yang, *Chem. Rev.*, 2012, **112**, 289.

125. B. G. Johnson, C. A. Gonzales, P. M. Gill and J. A. Pople, *Chem. Phys. Lett.*, 1994, **221**, 100–108.

126. B. S. Jursic, *Chem. Phys. Lett.*, 1996, **256**, 603–608.

127. G. I. Csonka and B. G. Johnson, *Theor. Chem. Acc.*, 1998, **99**, 158–165.

128. B. J. Lynch and D. G. Truhlar, *J. Phys. Chem. A*, 2001, **105**, 2936–2941.

129. S. Patchkovskii and T. Ziegler, *J. Chem. Phys.*, 2002, **116**, 7806.

130. E. R. Johnson, P. Mori-Sánchez, A. J. Cohen and W. Yang, *J. Chem. Phys.*, 2008, **129**, 204112.

131. A. J. Cohen, P. Mori-Sánchez and W. Yang, *Phys. Rev. B*, 2008, **77**, 115123.

132. P. Mori-Sánchez, A. J. Cohen and W. Yang, *Phys. Rev. Lett.*, 2008, **100**, 146401.

133. P. Mori-Sánchez, A. J. Cohen and W. Yang, *J. Chem. Phys.*, 2006, **125**, 201102.

134. A. Ruzsinszky, J. P. Perdew, G. I. Csonka, O. A. Vydrov and G. E. Scuseria, *J. Chem. Phys.*, 2007, **126**, 104102.

135. T. Bally and G. N. Sastry, *J. Phys. Chem. A*, 1997, **101**, 7923–7925.

136. Y. Xie, H. F. Schaefer, X.-Y. Fu and R.-Z. Liu, *J. Chem. Phys.*, 1999, **111**, 2532.

137. D. J. Tozer, *J. Chem. Phys.*, 2003, **119**, 12697–12699.

138. A. J. Cohen, P. Mori-Sánchez and W. Yang, *Science*, 2008, **321**, 792–794.

139. O. A. Vydrov, G. E. Scuseria and J. P. Perdew, *J. Chem. Phys.*, 2007, **126**, 154109.

140. A. Ruzsinszky, J. P. Perdew, G. I. Csonka, O. A. Vydrov and G. E. Scuseria, *J. Chem. Phys.*, 2006, **125**, 194112.

141. A. D. Dutoi and M. Head-Gordon, *Chem. Phys. Lett.*, 2006, **422**, 230–233.

142. R. Haunschild, T. M. Henderson, C. A. Jiménez-Hoyos and G. E. Scuseria, *J. Chem. Phys.*, 2010, **133**, 134116.

143. Y. Zhang and W. Yang, *J. Chem. Phys.*, 1998, **109**, 2604–2608.

144. J. P. Perdew and A. Zunger, *Phys. Rev. B*, 1981, **23**, 5048–5079.

145. S. Goedecker and C. J. Umrigar, *Phys. Rev. A*, 1997, **55**, 1765–1771.

146. O. A. Vydrov and G. E. Scuseria, *J. Chem. Phys.*, 2004, **121**, 8187–8193.

147. O. A. Vydrov and G. E. Scuseria, *J. Chem. Phys.*, 2005, **122**, 184107.

148. O. A. Vydrov, G. E. Scuseria, J. P. Perdew, A. Ruzsinszky and G. I. Csonka, *J. Chem. Phys.*, 2006, **124**, 094108.

149. A. M. Teale, F. De Proft and D. J. Tozer, *J. Chem. Phys.*, 2008, **129**, 044110.

150. J.-W. Song, M. A. Watson and K. Hirao, *J. Chem. Phys.*, 2009, **131**, 144108.

151. T. Tsuneda, J.-W. Song, S. Suzuki and K. Hirao, *J. Chem. Phys.*, 2010, **133**, 174101.

152. T. Körzdörfer, R. M. Parrish, J. S. Sears, C. D. Sherrill and J.-L. Bredas, *J. Chem. Phys.*, 2012, **137**, 124305.

153. M. Srebro and J. Autschbach, *J. Phys. Chem. Lett.*, 2012, **3**, 576–581.

154. T. Stein, J. Autschbach, N. Govind, L. Kronik and R. Baer, *J. Phys. Chem. Lett.*, 2012, **3**, 3740–3744.

155. A. M. Teale, F. De Proft, P. Geerlings and D. J. Tozer, *Phys. Chem. Chem. Phys.*, 2014, **16**, 14420–14434.

156. J. Autschbach and M. Srebro, *Acc. Chem. Res.*, 2014, **47**, 2592–602.

157. P. Mori-Sánchez and A. J. Cohen, *Phys. Chem. Chem. Phys.*, 2014, **16**, 14378–14387.

158. V. Vlček, H. R. Eisenberg, G. Steinle-Neumann, L. Kronik and R. Baer, *J. Chem. Phys.*, 2015, **142**, 034107.

159. S. R. Whittleton, X. A. Sosa Vazquez, C. M. Isborn and E. R. Johnson, *J. Chem. Phys.*, 2015, **142**, 184106.

160. M. J. G. Peach, A. M. Teale, T. Helgaker and D. J. Tozer, *J. Chem. Theory Comput.*, 2015, **11**, 5262–5268.

161. J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, V. N. Staroverov and J. Tao, *Phys. Rev. A*, 2007, **76**, 040501.

162. J. P. Perdew, R. G. Parr, M. Levy and J. L. Balduz, *Phys. Rev.*, 1982, **49**, 1691–1694.

163. Y. Zhang and W. Yang, *Theor. Chem. Acc.*, 1999, **103**, 346–348.

164. W. Yang, Y. Zhang and P. W. Ayers, *Phys. Rev. Lett.*, 2000, **84**, 5172–5175.

165. J. Perdew and M. Levy, *Phys. Rev. Lett.*, 1983, **51**, 1884–1887.

166. T. Koopmans, *Physica*, 1933, **1**, 104–113.

167. J. Janak, *Phys. Rev. B*, 1978, **18**, 7165–7168.

168. M. J. G. Peach, E. I. Tellgren, P. Sałek, T. Helgaker and D. J. Tozer, *J. Phys. Chem. A*, 2007, **111**, 11930–11935.

169. T. Heaton-Burgess and W. Yang, *J. Chem. Phys.*, 2010, **132**, 234113.

170. A. J. Johansson, M. R. A. Blomberg and P. E. M. Siegbahn, *J. Chem. Phys.*, 2008, **129**, 154301.

171. J. Poater, M. Solà, A. Rimola, L. Rodríguez-Santiago and M. Sodupe, *J. Phys. Chem. A*, 2004, **108**, 6072–6078.

172. R. Rios-Font, M. Sodupe, L. Rodríguez-Santiago and P. R. Taylor, *J. Phys. Chem. A*, 2010, **114**, 10857–10863.

173. T. Körzdörfer, S. Kümmel and M. Mundt, *J. Chem. Phys.*, 2008, **129**, 014110.

174. D. Hofmann and S. Kümmel, *Phys. Rev. B*, 2012, **86**, 201109.

175. X. Zheng, A. J. Cohen, P. Mori-Sánchez, X. Hu and W. Yang, *Phys. Rev. Lett.*, 2011, **107**, 026403.

176. X. Zheng, T. Zhou and W. Yang, *J. Chem. Phys.*, 2013, **138**, 174105.

177. C. Li, X. Zheng, A. J. Cohen, P. Mori-Sánchez and W. Yang, *Phys. Rev. Lett.*, 2015, **114**, 053001.

178. E. Kraisler and L. Kronik, *Phys. Rev. Lett.*, 2013, **110**, 126403.

179. O. A. Vydrov and G. E. Scuseria, *J. Chem. Phys.*, 2006, **125**, 234109.

180. É. Dumont, A. D. Laurent, X. Assfeld and D. Jacquemin, *Chem. Phys. Lett.*, 2011, **501**, 245–251.

181. J. Jaramillo, G. E. Scuseria and M. Ernzerhof, *J. Chem. Phys.*, 2003, **118**, 1068–1073.

182. J. P. Perdew, V. N. Staroverov, J. Tao and G. E. Scuseria, *Phys. Rev. A*, 2008, **78**, 052513.

183. A. V. Arbuznikov and M. Kaupp, *Chem. Phys. Lett.*, 2007, **440**, 160–168.

184. A. V. Arbuznikov and M. Kaupp, *J. Chem. Phys.*, 2014, **141**, 204101.

185. A. J. Cohen, P. Mori-Sánchez and W. Yang, *J. Chem. Phys.*, 2008, **129**, 121104.

186. P. Mori-Sánchez, A. J. Cohen and W. Yang, *Phys. Rev. Lett.*, 2009, **102**, 066403.

187. E. J. Baerends, *Phys. Rev. Lett.*, 2001, **87**, 133004.

188. M. Fuchs, Y.-M. Niquet, X. Gonze and K. Burke, *J. Chem. Phys.*, 2005, **122**, 094116.

189. M. J. G. Peach, A. M. Teale and D. J. Tozer, *J. Chem. Phys.*, 2007, **126**, 244104.

190. M. E. Casida, C. Jamorski, K. C. Casida and D. R. Salahub, *J. Chem. Phys.*, 1998, **108**, 4439–4449.

191. D. J. Tozer and N. C. Handy, *J. Chem. Phys.*, 1998, **109**, 10180–10189.

192. J. P. Perdew and K. Burke, *Int. J. Quantum Chem.*, 1996, **57**, 309–319.

193. D. J. Tozer and N. C. Handy, *J. Chem. Phys.*, 1998, **108**, 2545–2555.

194. M. E. Casida, K. C. Casida and D. R. Salahub, *Int. J. Quantum Chem.*, 1998, **70**, 933–941.

195. M. Grüning, O. V. Gritsenko, S. J. A. van Gisbergen and E. J. Baerends, *J. Chem. Phys.*, 2001, **114**, 652–660.

196. M. J. Allen and D. J. Tozer, *J. Chem. Phys.*, 2000, **113**, 5185–5192.

197. T. Yanai, R. J. Harrison and N. C. Handy, *Mol. Phys.*, 2005, **103**, 413–424.

198. M. J. G. Peach, T. Helgaker, P. Sałek, T. W. Keal, O. B. Lutnæs, D. J. Tozer and N. C. Handy, *Phys. Chem. Chem. Phys.*, 2006, **8**, 558–562.

199. J.-W. Song, T. Hirosawa, T. Tsuneda and K. Hirao, *J. Chem. Phys.*, 2007, **126**, 154105.

200. M. J. G. Peach, A. J. Cohen and D. J. Tozer, *Phys. Chem. Chem. Phys.*, 2006, **8**, 4543–4549.

201. R. Baer and D. Neuhauser, *Phys. Rev. Lett.*, 2005, **94**, 043002.

202. U. Salzner and R. Baer, *J. Chem. Phys.*, 2009, **131**, 231101.

203. T. Stein, H. Eisenberg, L. Kronik and R. Baer, *Phys. Rev. Lett.*, 2010, **105**, 266802.

204. T. Stein, L. Kronik and R. Baer, *J. Am. Chem. Soc.*, 2009, **131**, 2818–2820.

205. T. Stein, L. Kronik and R. Baer, *J. Chem. Phys.*, 2009, **131**, 244119.

206. N. Kuritz, T. Stein, R. Baer and L. Kronik, *J. Chem. Theory Comput.*, 2011, **7**, 2408–2415.

207. S. Refaely-Abramson, R. Baer and L. Kronik, *Phys. Rev. B*, 2011, **84**, 075144.

208. T. Minami, M. Nakano and F. Castet, *J. Phys. Chem. Lett.*, 2011, **2**, 1725–1730.

209. L. Kronik, T. Stein, S. Refaely-Abramson and R. Baer, *J. Chem. Theory Comput.*, 2012, **8**, 1515–1531.

210. A. Karolewski, L. Kronik and S. Kümmel, *J. Chem. Phys.*, 2013, **138**, 204115.

211. A. Karolewski, T. Stein, R. Baer and S. Kümmel, *J. Chem. Phys.*, 2011, **134**, 151101.

212. H. Sun and J. Autschbach, *ChemPhysChem*, 2013, **14**, 2450–2461.

213. P. Agrawal, A. Tkatchenko and L. Kronik, *J. Chem. Theory Comput.*, 2013, **9**, 3473–3478.

214. A. J. Garza, O. I. Osman, A. M. Asiri and G. E. Scuseria, *J. Phys. Chem. B*, 2015, **119**, 1202–1212.

215. E. Livshits and R. Baer, *Phys. Chem. Chem. Phys.*, 2007, **9**, 2932–41.

216. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09 Revision A.02*, 2009.

217. J.-D. Chai and M. Head-Gordon, *J. Chem. Phys.*, 2008, **128**, 84106.

218. J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615–6620.

219. C. Cardenas, P. Ayers, F. De Proft, D. J. Tozer and P. Geerlings, *Phys. Chem. Chem. Phys.*, 2011, **13**, 2285–2293.

220. R. D. Amos, I. L. Alberts, J. S. Andrews, A. J. Cohen, S. M. Colwell, N. C. Handy, D. Jayatilaka, P. J. Knowles, R. Kobayashi, G. J. Laming, A. M. Lee, P. E. Maslen, C. W. Murray, P. Palmieri, J. E. Rice, E. D. Simandiras, A. J. Stone, M.-D. Su and D. J. Tozer, Cadpac 6.5*, The Cambridge Analytic Derivatives Package*, 1998.

221. R. Baer, E. Livshits and U. Salzner, *Annu. Rev. Phys. Chem.*, 2010, **61**, 85–109.

222. M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. V. Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus and W. A. de Jong, *Comput. Phys. Commun.*, 2010, **181**, 1477–1489.

223. Values for $I_0^{N_0}$ and $A_0^{N_0}$ were obtained from the second and third columns of the table in the Appendix of Ref. 219, with the exception of the electron affinity of As, where a more recent value of 0.8048 eV was taken from Ref. 224, and the electron affinity of N, where the ground state value of zero was used.

224. W. M. Haynes and D. R. Lide, *CRC Handbook of Chemistry and Physics*, CRC Press, Boca Raton, 92nd edn., 2012.

225. L. Pauling, *J. Am. Chem. Soc.*, 1932, **481**, 3570–3582.

226. R. S. Mulliken, *J. Chem. Phys.*, 1934, **2**, 782–793.

227. W. Gordy, *Phys. Rev.*, 1946, **69**, 604–607.

228. W. Gordy, *J. Chem. Phys.*, 1951, **19**, 792.

229. A. D. Walsh, *Proc. R. Soc. Lond. A*, 1951, **207**, 13–22.

230. R. P. Iczkowski and J. L. Margrave, *J. Am. Chem. Soc.*, 1961, **83**, 3547–3551.

231. A. L. Allred and E. G. Rochow, *J. lnorg. Nucl. Chem*, 1958, **5**, 264–268.

232. R. T. Sanderson, *J. Am. Chem. Soc.*, 1983, **105**, 2259–2261.

233. L. C. Allen, *J. Am. Chem. Soc.*, 1989, **111**, 9003–9014.

234. R. G. Pearson, *Inorg. Chem.*, 1988, **27**, 734–740.

235. R. G. Pearson, *J. Am. Chem. Soc.*, 1985, **107**, 6801–6806.

236. R. G. Parr and W. T. Yang, *J. Am. Chem. Soc.*, 1984, **106**, 4049–4050.

237. P. W. Ayers and M. Levy, *Theor. Chem. Acc.*, 2000, **103**, 353–360.

238. K. Fukui, T. Yonezawa and H. Shingu, *J. Chem. Phys.*, 1952, **20**, 722.

239. S. Kato, *Theor. Chem. Acc.*, 2000, **103**, 219–220.

240. K. Fukui, *Science*, 1982, **218**, 747–754.

241. W. Yang, A. J. Cohen, F. De Proft and P. Geerlings, *J. Chem. Phys.*, 2012, **136**, 144110.

242. M. Levy and J. P. Perdew, *Phys. Rev. A*, 1985, **32**, 2010–2021.

243. S. Liu and R. G. Parr, *Phys. Rev. A*, 1996, **53**, 2211–2219.

244. A. Borgoo, A. M. Teale and D. J. Tozer, *J. Chem. Phys.*, 2012, **136**, 034101.

245. Q. Zhao, R. C. Morrison and R. G. Parr, *Phys. Rev. A*, 1994, **50**, 2138–2142.

246. S. Liu and R. G. Parr, *Phys. Rev. A*, 1997, **55**, 1792–1798.

247. R. C. Morrison and Q. Zhao, *Phys. Rev. A*, 1995, **51**, 1980–1984.

248. E. Fermi and E. Amaldi, *Accad. Ital. Rome*, 1934, **6**, 117–149.

249. P. W. Ayers, R. C. Morrison and R. G. Parr, *Mol. Phys.*, 2005, **103**, 2061–2072.

250. R. G. Parr and S. K. Ghosh, *Phys. Rev. A*, 1995, **51**, 3564–3570.

251. D. J. Tozer, N. C. Handy, R. D. Amos, J. A. Pople, R. H. Nobes, Y. Xie and H. F. Schaefer, *Mol. Phys.*, 1993, **79**, 777–793.

252. J. M. Galbraith and H. F. Schaefer III, *J. Chem. Phys.*, 1996, **105**, 862–864.

253. N. Rösch and S. B. Trickey, *J. Chem. Phys.*, 1997, **106**, 8940–8941.

254. N. C. Handy and D. J. Tozer, *Mol. Phys.*, 1998, **94**, 707–715.

255. A. J. Sadlej, *Theor. Chim. Acta.*, 1991, **79**, 123–140.

256. A. J. Sadlej, *Collect. Czech. Chem. Commun.*, 1988, **53**, 1995–2016.

257. T. W. Keal and D. J. Tozer, *J. Chem. Phys.*, 2003, **119**, 3015–3024.